## Claremont Colleges Scholarship @ Claremont

**CGU Theses & Dissertations** 

CGU Student Scholarship

Spring 2021

# Achieving Evaluation Influence Through Elaboration Likelihood Model-informed Evaluation Product Designs

Agnieszka Maria Helena Rykaczewska Claremont Graduate University

Follow this and additional works at: https://scholarship.claremont.edu/cgu\_etd

#### **Recommended Citation**

Rykaczewska, Agnieszka Maria Helena. (2021). *Achieving Evaluation Influence Through Elaboration Likelihood Model-informed Evaluation Product Designs*. CGU Theses & Dissertations, 225. https://scholarship.claremont.edu/cgu\_etd/225. doi: 10.5642/cguetd/225

This Open Access Dissertation is brought to you for free and open access by the CGU Student Scholarship at Scholarship @ Claremont. It has been accepted for inclusion in CGU Theses & Dissertations by an authorized administrator of Scholarship @ Claremont. For more information, please contact scholarship@cuc.claremont.edu.

Achieving Evaluation Influence Through Elaboration Likelihood Model-informed Evaluation

Product Designs

by

Agnieszka M.H. Rykaczewska

Claremont Graduate University

© Copyright Agnieszka M.H. Rykaczewska, 2021.

All rights reserved

#### **APPROVAL OF THE REVIEW COMMITTEE**

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

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

> > Stewart Donaldson

Claremont Graduate University

Distinguished University Professor

Leslie Fierro

Claremont Graduate University

Assistant Clinical Professor of Evaluation

Melvin Mark

Pennsylvania State University

Professor of Psychology

#### ABSTRACT

Achieving Evaluation Influence Through Elaboration Likelihood Model (ELM)-informed

**Evaluation Product Designs** 

Agnieszka M.H. Rykaczewska

Claremont Graduate University, 2021

The ultimate purpose of evaluation is social betterment, which is achieved through evaluation influence. Progress has been made in defining the mechanisms of evaluation influence (Mark & Henry, 2004); however, little research has explored how the design of evaluation products trigger these mechanisms. Sister fields such as persuasion psychology can provide guidance to fill this gap. The Elaboration Likelihood Model, a dual-processing model of persuasion, provides insights into how persuasive information is processed and how this processing impacts attitude formation and behavioral intention (Petty & Cacioppo, 1986). By translating the principles of the Elaboration Likelihood Model, this research explores how various data presentation conventions – minimalist, embellished, and interactive – impact evaluation influence.

In the first phase of this research, minimalist and embellished data visualization conventions did not result in differences in participant experience of the visualization nor different interpretation or attitudinal outcomes; however, motivation to elaborate significantly impacted both participant experiences and outcomes. Additionally, engagement with the data visualization played a role in how participants processed the evaluation findings, with highly engaged individuals basing their evaluand-specific attitudes on the strength of the evaluation findings. The second phase of this research demonstrated no significant differences in attitude strength and donation behaviors between minimalist and embellished data visualization. Instead, donation behaviors were driven by attitudes formed after reading the evaluation findings and motivation to elaborate. The final experiment found that interactive data presentations promoted elaboration and the formation of attitudes based on the strength of the evaluation findings. Additionally, significant differences in attitude persistence and behavioral intent were found based on the strength of evaluation findings; behavioral intent was additionally impacted by motivation to elaborate and engagement with the data presentation. Finally, donation behaviors were driven by motivation to elaborate, engagement with the data presentation, and evaluand-specific attitudes formed after reading the evaluation findings.

The results of this research demonstrate that the design of evaluation products and audience characteristics such as motivation to elaborate can be factors impacting evaluation influence. Based on these findings, evaluation practitioners can promote evaluation influence by seeking out opportunities to design products that increase audience involvement to support elaboration processes. The current research also identifies both risks to and opportunities for increased evaluation influence based on the audiences' level of motivation to elaborate, which provide guidance to evaluation practitioners seeking to maximize their evaluation's impact. More broadly, this research advances new directions for research on evaluation influence by providing empirical evidence for influence pathways, for data visualization research by demonstrating the importance of motivation to elaborate to visualization experience and outcomes, and for research on the application of Elaboration Likelihood Model principles within the context of evaluation.

#### ACKNOWLEDGEMENTS

I would like to thank the members of my dissertation committee for their support and invaluable feedback throughout my doctoral studies. I especially thank Dr. Tarek Azzam for his mentorship and guidance which were instrumental in my growth as a researcher and evaluator, and for his encouragement which helped me stay committed to my goal. I also greatly appreciate my CGU colleagues, in particular Blake Beckmann, Jeff Ramdas, Dana Wanzer, Kathleen Doll, and Charles Somerville, for their talents, discussions, and insights that strengthened the quality of this dissertation. I thank my parents, Anna and Krzysztof Rykaczewski, for inspiring me to merge art and science, and for providing emotional, mental, and financial support during my studies. I thank my brother and sister-in-law for sharing their wisdom and experience with me. And I thank my partner, Matthew Laney, for taking every single step of this journey with me – the impact of his support is beyond measure, even for an evaluator.

<b>TABLE OF</b>	<b>CONTENTS</b>
-----------------	-----------------

ABSTRACT	I
ACKNOWLEDGEMENTSV	1
LIST OF TABLES	<
LIST OF FIGURESX	/
CHAPTER 1 INTRODUCTION AND LITERATURE REVIEW	L
Overview of Evaluation Influence Theory	1
GAPS IN EVALUATION INFLUENCE THEORY	9
Addressing Gaps Through the Application of the Elaboration Likelihood Model	3
OTHER ROLES FOR EVALUATION IN THE ELABORATION CONTEXT	כ
RESEARCH QUESTIONS	7
Overview of Research Design	)
CHAPTER 2 PHASE I: EXPERIMENTAL EXPLORATION OF ELABORATION LIKELIHOOD ON PARTICIPANTS'	
EXPERIENCE AND OUTCOMES OF DATA VISUALIZATION42	2
METHODS45	5
RESULTS	3
Population	3
Analysis of Data Visualization Experience5٤	8
Analysis of Outcomes	4
Qualitative Analysis	2
Analysis of Influence Pathway82	1
DISCUSSION	7
CHAPTER 3 PHASE II: EXPERIMENTAL EXPLORATION OF EFFECTS OF ELABORATION AND DATA VISUALIZATION	
ON ATTITUDE STRENGTH AND BEHAVIOR97	7
METHODS	9

Recruitment	
RESULTS	102
Population	
Attitude Change Analyses	
Attitude Resilience Analyses	11(
Attitude Persistence Analyses	116
Behavioral Intent Analyses	
Donation Behavior Analyses	125
Influence Pathway Analyses	
DISCUSSION	130
BEHAVIOR	13
3EHAVIOR	13
METHODS	135
Recruitment	135
RESULTS	138
Engagement Analysis	
Attitude Change Analyses	144
Attitude Strength Analyses	151
Donation Behavior Analysis	
DISCUSSION	167
CHAPTER 5 DISCUSSION	173
INTERPRETATION OF RESULTS	173
LIMITATIONS	
THEORETICAL IMPLICATIONS	177
Evaluation Influence	
Data Visualization.	

Elaboration Likelihood Model.	
FUTURE RESEARCH	
PRACTICAL IMPLICATIONS	
Conclusions	
REFERENCES	
APPENDIX A. PREVIOUS OPERATIONALIZATIONS OF EVALUATION INFLUENCE	
APPENDIX B. RECRUITMENT AND DEBRIEF TEXTS	
APPENDIX C. SCENARIO TEXT	
APPENDIX D. MOTIVATION TO ELABORATE SCALE	
APPENDIX E. ATTITUDE SCALES	
APPENDIX F. TREATMENTS	
APPENDIX G. INTERPRETATION SURVEY ITEMS	
APPENDIX H. PERCEPTIONS SCALES	
APPENDIX I. DEMOGRAPHIC QUESTIONNAIRE	231
APPENDIX J. BEHAVIORAL INTENT SURVEY	
APPENDIX K. INTERACTIVE DATA DISPLAY	

### LIST OF TABLES

Table 1. Mark and Henry's (2004) model of evaluation influence. Source: Mark and Henry,
2004, p. 41
Table 2. MTurk demographics compared to US Population 46
Table 3. Phase 1 participants by age groups
Table 4. Participants' highest level of competed education. 55
Table 5. Initial sample participants' experience with data visualization (N=201).
Table 6. Motivation to elaborate scores by participant gender
Table 7. Summary of dependent variables related to data visualization experience
Table 8. Instances of broken multivariate normality assumption assessed by Shaprio-Wilk
statistic with Bonferroni correction
Table 9. Final sample size per condition for data visualization experience MANOVA analyses.
Table 10. Comparison of skew and kurtosis between original and transformed satisfaction
<i>scores</i>
Table 11. Summary of Dependent Variables Related to Data Visualization Experience    65
Table 12. Instances of broken multivariate normality assumption assessed by Shaprio-Wilk
statistic with Bonferroni correction (p<0.006)
Table 13. Final sample size per condition for outcome MANOVA analyses. 66
Table 14. Results of multivariate tests for 2-way interactions
Table 15. Crosstabulation of chart recognition and level of motivation to elaborate. 70
Table 16. Crosstabulation of chart recognition and data visualization approach
Table 17. Crosstabulation of chart recognition and evidence strength condition. 72

Table 18. Frequency and samples of qualitative codes for accuracy of interpretation.	74
Table 19. Crosstabulation of qualitative interpretation accuracy and evidence strength.	76
Table 20. <i>Example responses coded as positive judgements</i> $(n = 42)$ .	78
Table 21. Frequency and samples of unique follow-up qualitative codes ( $N = 120$ )	79
Table 22. Comparison of frequency of accuracy of initial and follow-up qualitative responses.	81
Table 23. Descriptive statistics for continuous variables included in regression analyses	82
Table 24. Moderation effects of evidence strength on engagement scores predicting global	
attitude change	83
Table 25. Descriptive statistics for evaluand-specific attitude change.	84
Table 26. Moderation effects of evidence strength on engagement scores predicting specific	
attitude change	85
Table 27. Gender demographics for Phase 2 research sample and subsamples	03
Table 28. Phase 2 participants by age groups. 1	04
Table 29. Participants' highest level of competed education. 1	05
Table 30. Descriptive statistics of attitude change and engagement scores for overall sample.	
	06
Table 31. Regression results predicting global attitude change	07
Table 32. Moderation effects of evidence strength on engagement scores predicting specific	
attitude change 1	09
Table 33. Descriptive statistics for dependent variables for attitude resilience three-way	
<i>MANOVA</i>	10
Table 34. Descriptive statistics for motivation to elaborate variable.    1	10

Table 35. Instances of broken multivariate normality assumption assessed by Shaprio-Wilk
statistic with Bonferroni correction ( $p < 0.006$ )
Table 36. Final sample sizes for resilience three-way MANOVA by condition
Table 37. Results of multivariate tests for two-way interactions. 113
Table 38. Descriptive statistics for resilience subsample data visualization engagement scores.
Table 39. Moderation effects of evidence strength on engagement scores predicting attitude
resilience
Table 40. Descriptive statistics for dependent variables for attitude persistence three-way
MANOVA
Table 41. Descriptive statistics for motivation to elaborate variable. 117
Table 42. Instances of broken multivariate normality assumption assessed by Shaprio-Wilk
statistic with Bonferroni correction (p<0.006)118
Table 43. Final sample size for persistence three-way MANOVA
Table 44. Results of multivariate tests for two-way interactions. 119
Table 45. Descriptive statistics of engagement scores for persistence subsample
Table 46. Moderation effects of evidence strength on engagement scores predicting attitude
persistence
Table 47. Descriptive statistics of engagement scores for overall Phase 2 sample. 123
Table 48. Moderation effects of evidence strength on engagement scores predicting behavioral
intent to support the evaluand
Table 49. Descriptive statistics for donation amount. 125

Table 50. Descriptive statistics for motivation to elaborate scores for donation behavior
analyses
Table 51. Instances of broken multivariate normality assumption assessed by Shaprio-Wilk
statistic with Bonferroni correction (p<0.006)126
Table 52. Comparison of three-way ANOVA findings using original and transformed donation
amount
Table 53. Descriptive Statistics for post intervention attitudes. 128
Table 54. Effects of post-intervention attitudes on predicting donation amount
Table 55. Gender demographics for Phase 3 research sample and subsamples
Table 56. Phase 3 participants by age groups
Table 57. Phase 3 participants' highest level of competed education
Table 58. Phase 3 Participants' Descriptive Statistics of Engagement Scores. 142
Table 59. Phase 3 Participants' descriptive statistics of motivation to elaborate scores.    142
Table 60. Frequency of categorical independent variables in three-way ANOVA assessing
engagement
Table 61. Instances of broken multivariate normality assumption assessed by Shaprio-Wilk
statistic with Bonferroni correction (p<0.006)143
Table 62. Phase 3 participants' dependent and independent variable descriptive statistics 145
Table 63. Moderation effects of evidence strength on presentation approach predicting global
attitude change
Table 64. Descriptive statistics for difference between expected and actual evaluation findings.

Table 65. Moderation effects of evidence strength on presentation approach predicting
evaluand-specific attitude change150
Table 66. Phase 3 descriptive statistics for attitude strength dependent variables.    152
Table 67. Phase 3 resilience subsample descriptive statistics of independent variables.    152
Table 68. Frequency of conditions in resilience subsample. 153
Table 69. Regression model predicting evaluand-specific attitude resilience.    154
Table 70. Descriptive statistics of independent variables predicting global attitude persistence.
Table 71. Frequency of conditions in persistence subsample
Table 72. Final regression model predicting global attitude persistence. 156
Table 73. Final regression model predicting evaluand-specific attitude persistence.    159
Table 74 Final regression model predicting behavioral intent    161
Table 75. Phase 3 participants' donation amount descriptive statistics. 163

## LIST OF FIGURES

Figure 1. Hypothetical elaboration pathways to evaluation influence.	14
Figure 2. Petty and Cacioppo's (1986) Elaboration Likelihood Model. Source: Petty &	
Сасіорро, 1986, р. 126	16
Figure 3. Elaboration Likelihood Model supplementing Ajzen's (1991) Theory of Planned	
Behavior	21
Figure 4. Hypothesized pathways from data visualization design to attitude change and dor	nation
behaviors	39
Figure 5. Relationships between the three research phases	40
Figure 6. Hypothesized evaluation influence pathway explored in Phase 1	44
Figure 7. Seven stages of Phase 1 Study	47
Figure 8. Mean duration examining visual by level of motivation to elaborate and evidence	
strength	61
Figure 9. Mean engagement score by level of motivation to elaborate.	62
Figure 10. Statistically significant differences in satisfaction with visual based on motivation	on to
elaborate	64
Figure 11. Mean trustworthiness scores based on level of motivation to elaborate and evide	ence
strength	68
Figure 12. Differences in accuracy of interpretation scores by level of motivation to elaborate	ate.
	69
Figure 13. Differences in attitude change specific to evaluand by level of motivation to	
elaborate	69

Figure 14. Predicted global	l attitude change based on centered engagement scor	es and evidence
strength		
Figure 15. Predicted evalue	and-specific attitude change based on centered engag	ement scores
and evidence strength		86
Figure 16. Eight stages of H	Phase 3 Procedure	100
Figure 17. Effects of engage	ement on predicted global attitude change	107
Figure 18. Effects of eviden	nce strength on predicted global attitude change	108
Figure 19. Predicted evaluation	and-specific attitude change based on centered engag	ement scores
and evidence strength		109
Figure 20. Average behavio	oral intent scores based on level of motivation to elabo	orate 113
Figure 21. Average global a	attitude change scores based on evidence strength	
Figure 22. Mean behaviora	al intent to support the evaluand score by motivation t	o elaborate
level		120
Figure 23. Mean Global At	titude Change score by motivation to elaborate level.	121
Figure 24. Predicted behav	vioral intent based on centered engagement scores and	d evidence
strength		125
Figure 25. Effect of global p	post-intervention attitudes on donation behavior	129
Figure 26. Effect of evaluan	nd-specific post-intervention attitudes on donation beh	<i>avior.</i> 130
Figure 27. Average engage	ment scores based on motivation to elaborate	
Figure 28. Interaction effect	ct between evidence strength and presentation approa	ch on predicted
global attitude change, hola	ling motivation to elaborate, engagement scores, surp	orise scores
constant		

Figure 29. Interaction effect between evidence strength and presentation approach on predicted
evaluand-specific attitude change, holding motivation to elaborate, engagement scores, surprise
scores constant
Figure 30. Effect of evidence strength on predicted global attitude persistence scores, holding
motivation to elaborate scores, engagement scores, surprise scores, interactive presentation, and
strong evidence constant
Figure 31. Effect of evidence strength on predicted evaluand-specific attitude persistence,
holding motivation to elaborate scores, engagement scores, surprise scores, and interactive
presentation constant
Figure 32. Effect of motivation to elaborate on predicted behavioral intent scores, holding
engagement scores, surprise scores, interactive presentation, and strong evidence constant 162
Figure 33. Effect of engagement scores on predicted behavioral intent scores, holding of
motivation to elaborate, surprise scores, interactive presentation, and strong evidence constant.
Figure 34. Effect of evidence strength on predicted behavioral intent scores, holding motivation
to elaborate scores, engagement scores, surprise scores, and interactive presentation constant.
Figure 35. Effect of engagement on predicted donation amounts, holding motivation to
elaborate, surprise scores, post-intervention attitudes, interactive visualization with strong
evidence constant
Figure 36. Effect of post-intervention evaluand-specific attitudes on predicted donation
amounts, holding motivation to elaborate scores, surprise scores, engagement scores, interactive
visualization with strong evidence constant

Figure 37. Effect of motivation to elaborate on predicted donation amounts, holding post-	
intervention evaluand-specific attitudes, surprise scores, engagement scores, interactive	
visualization with strong evidence constant1	67

#### **CHAPTER 1**

#### INTRODUCTION AND LITERATURE REVIEW

The ultimate purpose of evaluation is to achieve social betterment through the utilization of evaluation findings and recommendations to inform policies and programs (Henry, 2000). However, evaluators often have limited power in how evaluative information is used. This observation has led to a shift in focus from instrumental use to evaluation influence (Kirkhart, 2000; Alkin & King, 2017; Mark & Henry, 2004). Proponents argue that instrumental use – or the use of evaluation findings for decision-making – is too limiting and that evaluation influence is a broader term than utilization that better captures the full, often complex, story of an evaluation's impact (Weiss, 1998; Kirkhart, 2000). Though much progress has been made in building a theory of evaluation influence, there remains a need for further definition and operationalization of its components and translation into practice.

Evaluation influence is "the capacity or power of [evaluation] to produce effects on others by intangible or indirect means" and serves as the link between evaluation products and evaluation use leading to social betterment (Kirkhart, 2000, p. 7; Mark & Henry, 2004). Mark and Henry (2004) identified four mechanisms of evaluation influence: general influence, cognitive and affective, motivational, and behavioral. Additionally, these mechanisms can occur at three levels: individual, interpersonal, and communal. Mark and Henry argue that evaluation processes and products trigger these underlying mechanisms to produce evaluation influence. Influence at one level can then trigger influence at a different level, creating influence pathways.

While framing evaluation influence in this manner creates a more dynamic model with several conceptual benefits, the utility of the Mark and Henry (2004) framework is still limited by the lack of operationalization of the four mechanisms (Herbert, 2014). A review of research

on evaluation influence revealed that this lack of operationalization has resulted in difficulty empirically establishing the presence of influence mechanisms. Additionally, though theoretically linked, it is unclear how these mechanisms manifest through evaluation products, leading to difficulties in the intentional application of the framework (Herbert, 2014). Practitioners face similar limitations from the framework. While they may be able to track influence pathways after-the-fact, those seeking to intentionally design evaluation products to promote influence have no guide to do so.

For example, data visualization is a common component of evaluation products, yet there is no guidance on whether data visualization can promote evaluation influence, and if so, which design features do so best. There is a wide variety of data visualization options and design conventions available for use by evaluators, and these can promote different effects. High data-ink visualizations, characterized by high proportions of ink used in a visualization dedicated to representing data, tend to promote faster and more accurate interpretation of data while more embellished graphs have been found to be more engaging and memorable (Bateman et al., 2010). Bateman and colleagues (2010) argued that embellished graphs could be more memorable for three reasons: (1) the presence of an image provides additional encoding in memory which can improve recall, (2) the embellished images are very different from one another, while plain bar charts all have similar appearances, and (3) viewer's emotional response to the imagery in embellished charts could help anchor the chart details in the viewer's memory.

Given that evaluation influence pathways likely take time, *engagement* and *memorability* may play key roles. Further exploration is needed to determine which data visualization design approach best supports evaluation influence, for whom, and under what circumstances. Thus,

while evaluation influence may be conceptually clearer than evaluation use, it still lacks sufficient clarity to be directly translated into practice.

Evaluation's sister fields – such as psychology, communications, and sociology – may provide valuable insights into the design of evaluation products for evaluation influence. As noted by Mark and Henry (2004), the Elaboration Likelihood Model (ELM), a dual-processing model of persuasion, may shed light on the general influence process by which evaluation products trigger the cognitive and affective mechanism of evaluation influence (Petty & Cacioppo, 1986). The ELM defines two pathways to attitude change, the central and peripheral routes, and the circumstances under which each path is likely to occur.

The central route is characterized by deep reflection on and assessment of the arguments presented in a persuasive message to determine attitude while the peripheral route does not process this information at all. Instead, those that take the peripheral route rely on pre-established rules (e.g., "I'll never agree to abortion") or other cues (e.g., an attractive message source) to determine their attitude. Research has demonstrated that attitudes changed via the central route tend to be more stable, persistent, and predictive of behaviors than those changed through the peripheral route (Petty & Cacioppo, 1986). This suggests that evaluation products seeking to trigger this general influence mechanism to create long-term influence pathways may be more effective if designed in a manner that encourages stakeholders towards central route processing of the evaluation's conclusions and evidence.

This research explores how the Petty and Cacioppo's (1986) Elaboration Likelihood Model can inform the design of evaluation products that trigger Mark and Henry's (2004) cognitive and affective mechanism of evaluation influence. This first chapter provides an overview of current literature on evaluation influence, using current debates in data visualization

design as an example to highlight challenges in applying the theory. Next, the paper explores how the ELM can be used to augment evaluation influence theory, address gaps, and provide valuable insights into design decisions. Next, chapters II, III, and IV summarize results from three phases of experimental research which explore how level of motivation to elaborate, data visualization approach, and evidence strength impact experience of data visualizations, attitudes, and pro-social behaviors. The final chapter provides discussion of findings, limitations, and directions for future research.

#### **Overview of Evaluation Influence Theory**

The concept of evaluation influence was born from a need to understand the impact of evaluation and the limitations of evaluation use theories to capture these impacts. As Henry (2000) re-established, the ultimate purpose of evaluation is to lead to social betterment. By identifying what works, for whom, under what circumstances, and why, evaluation can serve as a powerful tool to inform decision-makers and improve social programs (Gargani & Donaldson, 2011). However, evaluation also requires significant resources and effort, and it is important to consider whether this investment will produce enough impact to be judged worthwhile. Traditionally, this conversation has been framed largely in terms of evaluation use (Kirkhart, 2000). To better judge the extent to which an evaluation has been used, theorists have identified a variety of types of uses, including symbolic, conceptual, enlightenment, process, instrumental, and misuse (Kirkhart, 2000; Patton, 1998; Shulha & Cousins, 1997; Weiss, 1998).

Recently, however, several issues have been raised regarding the focus on evaluation use. First, despite the longstanding focus on use in evaluation literature, it is difficult to attain in practice. Fleischer and Christie (2009) found that 68% of American Evaluation Association members surveyed reported non-use of evaluation results. Results also indicated that barriers to

evaluation use include not only evaluation factors, but also human and contextual factors. In short, whether an evaluation's findings are directly utilized is often beyond the evaluator's control.

Despite these findings, or perhaps because of, evaluation use has continued to be emphasized in theory, research, and practice, to levels that some have called inappropriate. Henry (2000) warned that the over-focus on use has led to goal displacement in the field. Rather than being a means to an end, Henry (2000) observed that use, rather than social betterment, had become the metric by which an evaluation's impact was judged.

Several issues have also been raised with evaluation use as a concept. Kirkhart (2000) argued that evaluation use lacked conceptual clarity such that the various types of use often overlapped and were vaguely defined. Enlightenment use, for example, could also be categorized as instrumental or conceptual use. Furthermore, Kirkhart (2000) argued that the concept of use was too narrow, focusing only on intended, results-based impact rather than capturing the full spectrum of an evaluation's impact. Mark and Henry (2004) expanded on these criticisms, arguing that "categories of use are distinguished by qualitatively different attributes," sometimes being defined by the type of change (conceptual and instrumental use), the intent (symbolic use), or by the source of use (process use)" (Mark and Henry, 2004, p. 36).

Together, these criticisms have caused some to shift their focus from use to evaluation influence. Influence, or the ability of an evaluation to produce effects on others, addresses the shortcomings outlined above. It accounts for the limited ability of the evaluator to directly enforce the use of evaluation findings and acknowledges that evaluation is often several steps removed from direct services. As a result, evaluation can only achieve social betterment by influencing others to change – be it in their thinking, their actions, or their programs. Similarly,

as argued by Mark and Henry (2004), the term "influence" acknowledges that an evaluation's impact is rarely a straight link but instead follows complex pathways involving many players.

The term influence may also reorient the focus of the field back towards social betterment. Rather than being an end-state, "influence" suggests that it is a conversation starter, pushing others towards a bigger end goal. Though the difference may seem simple semantics, several evaluation theorists have argued that language used to describe evaluation shapes our understanding and perceptions of the field and ultimately guides our actions (Kirkhart, 2000; Patton, 2000). Thus, resolving nuanced differences in terminologies is critical for ensuring that the correct perspective – in this case social betterment as the ultimate goal of evaluation – is maintained.

The shift towards evaluation influence has also resulted in models that are simultaneously conceptually clearer and better able to capture complexity than the taxonomies of evaluation use. By building on Kirkhart's (2000) conceptual work and integrating research from the social sciences, Mark and Henry (2004) proposed a framework for evaluation influence. The model is organized by four types of influence mechanisms (general influence, cognitive and affective, motivational, and behavioral) that can occur at any of three levels of analysis (individual, interpersonal, and collective; See Table 1).

The four types of influence mechanisms in Mark and Henry's (2004) framework are descriptive of the type of underlying change that occurs when an evaluation product or process exerts influence. General influence describes the fundamental structures of change, such as priming, persuasion, and standard setting. General influence mechanisms are often supports for the remaining mechanisms. Cognitive and affective changes are those that alter ways of thinking

and feeling. For example, an evaluation report on a youth program can make social issues faced by youth more salient to readers of the report.

#### Table 1.

Mark and Henry's (2004) model of evaluation influence. Source: Mark and Henry, 2004, p. 41.

Type of Process/Outcome	Level of Analysis			
	Individual	Interpersonal	Collective	
General Influence	Elaboration	Justification	Ritualism	
	Heuristics	Persuasion	Legislative hearings	
	Priming	Change agent	Coalition formation	
	Skill acquisition	Minority-opinion	Drafting legislation	
		influence	Standard setting	
			Policy consideration	
Cognitive and affective	Salience Opinion/attitude valence	Local descriptive norms	Agenda setting Policy-oriented learning	
Motivational	Personal goals & aspirations	Injunctive norms Social reward Exchange	Structural incentives Market forces	
Behavioral	New skill performance Individual change in practice	Collaborative change in practice	Program continuation, cessation, or change Policy change Diffusion	

While at the individual level, cognitive and affective processes are typically captured by attitude change, at the interpersonal level these changes are better represented as shifts in group norms. At the collective level, cognitive and affective mechanisms are inclusive of agenda setting and policy-oriented learning. Motivational mechanisms, on the other hand, represent shifts in goals and aspirations. As Mark and Henry (2004) outline, even the simple presence of an ongoing evaluation may shift motivation towards higher levels of performance. At the collective level, motivational mechanisms of evaluation influence are ones where evaluation processes and products affect change in actions, including changes in programmatic practices and changes in funding.

Using these components as links in pathways of influence allows the Mark and Henry (2004) framework to fully capture the complex ways in which evaluation impact occurs. That is, evaluation influence is simultaneously an outcome and a process such that each mechanism represents an impact obtained and could also be a trigger for the next mechanism in the pathway. For example, a single individual reading an evaluation report may then become a change agent that successfully advocates for shifts in incentive structures and policy changes. In this scenario, the original cognitive and affective mechanism then triggered a general influence mechanism which in turn led to motivational and behavioral changes at the collective level. Links in pathways of influence, Mark and Henry (2004) argue, do not have to be linear, but often can circle back until the full impact of the evaluation is captured.

Mark and Henry (2004) suggest several ways their framework improves on previous conceptualizations of evaluation influence. Primary among these is the specification of influence mechanisms, which not only bring additional conceptual clarity about the nature of evaluation influence, but also provide the ability to make "concrete predictions about the general relations between different components" of the model (Mark & Henry, 2004, p. 47). In short, by specifying the mechanisms, Mark and Henry (2004) bring greater understanding as to how evaluation influence fits into the larger picture of an evaluation's impact. Similarly, the influence mechanisms make it easier to trace complex pathways of influence, acknowledging that the route to impact is rarely direct. Finally, the inclusion of levels of analysis in the framework raise awareness of important contextual factors to consider when assessing evaluation impact (Mark & Henry, 2004).

The evaluation influence literature has implications for researchers and practitioners alike. Both Kirkhart (2000) and Mark and Henry (2004) have expanded discussions of impact

beyond the traditional evaluation use. For researchers, this means additional outcomes to consider when conducting empirical research on evaluation. Though Fleischer and Christie (2009) found scant evidence of impact when focused on evaluation use, broadening the scope to evaluation influence may be better for empirically documenting the impact that evaluations may have. Additionally, the specifications of influence mechanisms and levels of analysis add further complexity to these models, and may enable researchers to better identify for whom, how, and under what circumstances evaluations are impactful. However, this additional complexity can also create challenges for measuring an evaluation's influence, especially when influence pathways are not linear.

Impact pathways may be particularly important for evaluation practitioners to consider. As noted by surveyed evaluators in Fleischer and Christie's (2009) study, part of what makes direct evaluation use so difficult to attain is that evaluators rarely have direct control over programs or policies and face contextual barriers to instrumental use. Evaluation influence literature provides these practitioners with a broader range of options and may allow them to better identify alternative pathways when faced with barriers. Additionally, practitioners may be better able to track their impact using the evaluation influence frameworks. Rather than waiting to the end of an evaluation to assess impact, practitioners may be able to utilize Mark and Henry's (2004) influence mechanisms as a guiding framework for ongoing tracking. This may allow for quicker identification of barriers and faster course corrections should an evaluation not be making sufficient impact.

#### **Gaps in Evaluation Influence Theory**

While the shift from use to evaluation influence has carried several benefits for researchers and practitioners alike, current models of evaluation influence are not without

shortcomings that could benefit from further development. The most significant gap, identified by Herbert (2014) in a review of research on evaluation influence, is a lack of operationalization which has hampered the application of the theory in practice. Herbert (2014) found that almost all published literature applying either Kirkhart's (2000) model or Mark and Henry's (2004) framework - from case studies to empirical research - lacked detail as to how evaluation influence or its mechanisms were identified. The case studies revealed that the lack of operationalization resulted in challenges for practitioners wishing to track the impact of their work. For researchers, it resulted in difficulties creating and implementing measures of evaluation influence.

The challenge in operationalizing evaluation influence stems from both the framework's structure and its complexity. In its current iteration, the mechanisms do not represent processes but rather categories. By themselves, they lack the detailed definition necessary for identification and application of these mechanisms in practice. This limitation is partially addressed by the listing of example processes, such as elaboration, agenda setting, and policy change, that fit within each category. Each of these, however, has its own extensive literature base built from substantial research in areas outside of evaluation. While standing on the shoulders of these giants certainly has advantages, it requires a level of expertise in these fields that may not be present in the typical evaluation researcher or practitioner, limiting the utility of the framework. Even when the expertise is present, it takes an additional step to translate principles and lessons learned from general research to the evaluation context.

Encouragingly, some progress towards bridging this gap has recently been made. As part of a larger work to trace evaluation influence within international humanitarian aid agencies, Oliver (2008) defined and codified a total of 15 phenomena using Mark and Henry's (2004)

model as a guiding framework (See Appendix A, Table A1). Though the codes were quite specific to the context and purpose of the study, Oliver (2008) provides an example of how the Mark and Henry (2004) framework may be further defined in a way that makes it easier to apply in practice.

Oliver (2008) provides the first steps in unpacking the evaluation influence mechanisms by clarifying what these processes look like within an evaluation context. Oliver's definitions should facilitate the identification of influence mechanisms for both researchers and practitioners. However, they are limited in their ability to guide intentional application of the mechanisms because they fail to capture how these states are generated. In short, they define the evaluation influence mechanisms as outcomes but not as processes, and as such, capture only a part of Mark and Henry's (2004) framework.

The implication of this gap is that evaluation researchers and practitioners know what effect they wish to create, but not how to do so. A concrete example of this challenge is the effective use of data visualization in evaluation products. Advocates of the use of data visualization in evaluation argue that, among other advantages, it facilitates communication of evaluation findings (Azzam, Evergreen, Germuth, & Kistler, 2013). This in turn may play a role in many of the influence mechanisms. However, as shown by Brown and Newman (1982), different data visualization presentations effect the degree to which stakeholders agree with evaluation recommendations and their consequent decision-making.

Brown and Newman presented teachers and school administrators with simulated evaluation reports that contained either no data, percentage data only, percentage data and graphs, or percentage data, graphs, and inferential statistics, and asked participants to rate the extent of agreement with recommendations, ratings of usefulness of the information, and ratings

of the evaluator. Findings indicate that participants were most likely to be swayed by the information when it was presented as percentage data and graphs. Conversely, participants showed least agreement with recommendations when the reports included percentage data, graphs, and inferential statistics. Similar trends were found in ratings of usefulness of the information and of the evaluator. Though limited in the types of data visualization explored, such findings highlight that presentation style and visualizations can affect the extent of an evaluation report's influence. How, then, should data visualization be designed so it best promotes evaluation influence?

This question may be central to several debates amongst researchers of data visualization. One such debate focuses on minimalist versus embellished graphs. Minimalist graphs are characterized by high proportions of ink used in a visualization dedicated to representing data, known as high data-ink ratios, and the elimination of any unnecessary features of a graph, referred to as "chart junk" (Tufte, 1983). Based on principles of human perception, advocates of minimalist visualizations argue that these promote fast and accurate understanding of the data displayed. This approach has been widely held as the standard for data visualization for the past few decades, including in the field of evaluation (Evergreen & Metzner, 2013). However, more recently, research on embellished graphs, which feature use of symbolism and often low proportions of ink used to represent data (low data-ink ratios), has advocated that such graphs are more engaging and memorable than the minimalist approach (Bateman et al., 2010; Byrne, Angus, & Wiles, 2016). Given that Mark and Henry (2004) note that evaluation pathways often take time to form, memorability may play an important role in promoting such pathways. Therefore, it is possible that the data visualization design approach utilized in an evaluation product may make evaluation influence more or less likely to occur.

Similar debates have been held over the use of interactive graphs, which utilize technological advances that turn the audience from "a passive consumer into an active participant" (Weissgerber et al., 2017, p. 20592). Using speed and accuracy of data interpretation as the standards for effectiveness, research has largely judged interactive and animated graphs as less effective than static graphs (Fisher, 2010). However, advocates of interactive graphs note that such visualization methods have advantages that may outweigh these risks. For example, Weissgerber et al. (2016) argue that interactive graphing allows for greater transparency by allowing for the exploration of more nuanced data patterns. The authors argue that by allowing for close examination of such components as standard deviation and error bars, features which typically would either overlap or be dropped in static graphs, audience members gain a more accurate understanding of the extent of difference between groups and limitations of the data. Additionally, cognitive research suggests that communication which includes active participation may be more effective than passive presentation of information (Natter & Berry, 2005). Therefore, it is unclear whether it may be advantageous to use interactive graphs as a component of evaluation products.

#### Addressing Gaps Through the Application of the Elaboration Likelihood Model

As these examples highlight, the link between evaluation products and evaluation influence needs further definition to help guide product design decisions. While research has yet to be completed to test these relationships, sister fields such as psychology can provide valuable insights and promising directions through the translation of general principles and theories of behavior to the evaluation context (Fleming, 2011). For example, Mark and Henry (2004) note that elaboration is one example of a general influence process which can trigger the cognitive and affective mechanism at the individual level. First defined by Petty and Cacioppo in 1981, elaboration is "the extent to which a person thinks about the issue-relevant arguments contained in a message" (Petty & Cacioppo, 1986, p. 128). Based on extensive research, Petty and Cacioppo proposed the Elaboration Likelihood Model (ELM), a dual processing model of persuasion that specifies when elaboration is and is not likely to occur. The ELM is a particularly promising theory because of its large body of supporting research, which has explored the theory's complexities with sufficient nuance to provide valuable lessons that can be applied across evaluation's multiple contexts and contingencies. Applying and translating principles of the ELM to an evaluation-specific context can help clarify which evaluation product designs, under what circumstances, are most likely to produce elaboration and the evaluation influence that results from it. Figure 1 outlines how such a process may connect to evaluation influence.

#### Figure 1.

Hypothetical elaboration pathways to evaluation influence.<sup>1</sup>



<sup>&</sup>lt;sup>1</sup> From Mark and Henry, 2004

The key feature of the ELM is its dual-processing design which delineates two pathways by which an individual's attitude towards an issue is determined (Petty & Cacioppo, 1986). The starting point of both routes is the presentation of a persuasive message which contains arguments for a certain stance on an issue. The central route is defined by deep consideration of the persuasive message's arguments (elaboration). If an individual perceives the arguments to be strong and valid, their attitudes towards the issue shift accordingly. Weak arguments, on the other hand, do not shift attitudes. Research has shown that attitudes changed by the central route have better persistence, resilience, and are more predictive of behaviors, collectively referred to as attitude strength (e.g., Crano, 1995; Haugtvedt, Schumann, Schneier, and Warren, 1994; Petty & Cacioppo, 1986; Verplanken, 1991). However, the central route relies on high levels of motivation and ability to elaborate on the message's content. When these are not present, individuals tend to take the peripheral route. The peripheral route is defined by a lack of elaboration and reliance on cues and simple rules to determine attitudes towards an issue (See Figure 2). For the peripheral route, the information contained in the message is of no importance. Instead, the attitude determination is made based on factors such as how the message was delivered or shaped (Petty & Cacioppo, 1986).

#### Figure 2.

*Petty and Cacioppo's (1986) Elaboration Likelihood Model. Source: Petty & Cacioppo, 1986, p. 126* 



This brief model overview suggests three initial implications for the design of evaluation products that support elaboration and evaluation influence. First, it suggests a certain structure for the presentation of evaluative conclusions and their evidence. Second, the importance of central route processing to attitude strength implies that evaluation products which promote such processing may be more effective at producing evaluation influence that results in influence pathways. Finally, it identifies two factors - motivation and ability to elaborate - that promote central route processing which may be used to guide product designs. Each of these implications will be explored further in the following sections.

All persuasive messaging has the same basic structural components – a stance supported by persuasive arguments – that is needed to engage an individual in attitude change (Petty & Cacioppo, 1986). Areni (2003) summarized two basic structural models for persuasive communication: the syllogistic model and jurisprudence model. Common components to both models was a conclusion or claim, as well as a series of premises or data that support the conclusion or claim. In the syllogistic model, strong messages were those that contained true premises which were connected to the conclusion in a logical and valid manner.

In the jurisprudence model, a claim that "includes qualifiers and rebuttals to alleviate counter-argumentation" resulted in stronger, more persuasive messages (Areni, 2003, p. 369). This implies that it is important for evaluation products to articulate a conclusion or claim which is supported through persuasive arguments based on the data. Reframing evaluation as a type of persuasive communication is not a new concept, as Newman, Brown, and Braskamp (1980) have already argued for this. At the core, program evaluation assigns a value-judgement based on systematically gathered evidence (Shadish, Cook, & Leviton, 1991). This is reflective of the structure of a persuasive message – the stance (the value-judgement or recommendation) which is supported by arguments (the systematically gathered evidence).
Defining the structure of evaluative communication carries significant implications for data visualization design. Specifically, it underscores the importance of clearly communicating the value-judgement or evaluative recommendations. Bateman et al. (2010) found that those who viewed embellished graphs were better able to discern the value message these visualizations contained. Though minimalist data visualization is lauded as objective because "by removing embellishment and non-data ink...the data 'is allowed to speak for itself", Bateman and colleagues argue that the process underlying data visualization inherently is biased, and embellished graphs merely make this already existing bias more explicit (Bateman et al., 2010, para. 73). In this way, embellished graphs may better reflect the structure of persuasive messaging by more effectively communicating a stance, thereby facilitating the elaboration process.

The second implication for evaluation product design from the ELM is that products which promote central route processing may result in attitudes which are more likely to lead to evaluation influence pathways. Research has shown that attitudes changed by the central route has better persistence, resilience, and is more predictive of behaviors, collectively referred to as attitude strength (e.g., Crano, 1995; Haugtvedt, Schumann, Schneier, and Warren, 1994; Petty & Cacioppo, 1986; Verplanken, 1991). Attitude persistence refers to the maintenance of an attitude over longer periods of time. Early work by Petty, Cacioppo, and Heesacker (1985, as cited in Petty & Cacioppo, 1986) manipulated likelihood of elaboration through varying factors related to motivation to elaborate and showed that two weeks later only high relevance participants maintained their position while those in the low relevance condition reverted to their original attitudes.

Similarly, in a field study of attitudes towards large scale use of coal for electricity, Verplanken (1991) measured elaboration likelihood by personal factors linked to motivation to elaborate. Findings indicate that attitude change among those with high elaboration likelihood persisted when re-measured 14 months later, but not so among those with low elaboration likelihood. In each of these studies, those who attained attitude change via the central route tended to have more stable attitudes than those who did so via the peripheral route. Given that evaluation influence pathways are likely to take time, persistent attitudes achieved through central route processing may enable the formation of these pathways.

Attitude resilience refers to the ability to resist counter-persuasion once initial attitude change is obtained and has also been shown to be affected differentially by the two routes of attitude change. For example, Haugtvedt, Schumann, Schneier, and Warren (1994) showed that individuals who based their attitudes on repeated advertisements with substantive variation showed similarly positive attitudes towards the product as those who were presented repeated advertisements with cosmetic variations. However, when participants were later presented with unfavorable information about the product, only those who formed their opinions based on advertisements with cosmetic variations changed their views. Meanwhile, those who formed their opinions on substantive variants maintained their positive views of the product. Within the evaluation context, stakeholders often seek multiple sources of input and information in addition to evaluations to guide their decision-making (Kirkhart, 2000). As such, evaluation products that promote the formation of resilient attitudes may be more effective at increasing the influence power of evaluation relative to the other sources of information stakeholders may encounter.

The final component of attitude strength is behavioral predictivity, which is also better supported through central route processing. Crano and Sivacek (1982, as cited in Crano, 1995)

explored how attitudes based on vested interest, a motivational factor which promotes central route processing, affected behaviors. The study took advantage of a proposed law to increase the drinking age for alcohol. Crano and Sivacek recruited university students who were against the law and categorized them as either high, moderate, or low vested interest based on their age, such that younger students who would be affected by the outcome of the proposed law were categorized as high vested interest and most likely to engage in central route processing. Though all recruited students were against the law, the study found nearly half of those with highly vested interest agreed to join an anti-referendum campaign while only 12% of those in the low vested interest did so. This aligns with predictions made by Petty and Cacioppo (1986) that those with higher involvement would be more likely to form attitudes via the central route, and that these attitudes would then be more predictive of behavior.

The behavioral predictivity aspect of attitude strength has the greatest implications for the formation of evaluation influence pathways. For one, increasing attitude strength maximizes the chances to move influence from a general or cognitive type of process to the behavioral process in Mark and Henry's (2004) model (see Table 1). However, other movements across the evaluation influence model require behavioral changes. For example, Mark and Henry place "change agent" as a general influence process on the interpersonal level. However, change agent also requires certain actions, such as advocating and promoting a cause. Therefore, movement at various points in Mark and Henry's model requires some degree of behavioral change and may therefore be best supported through central route processing of evaluation products.

It should be noted, however, that central route processing doesn't guarantee that individuals will act on their changed attitudes, but it does increase the likelihood that they will follow through. The Theory of Planned Behavior notes three factors that affect the likelihood of

a behavior: (1) attitudes towards the behavior, (2) subjective norms regarding the behavior, and (3) perceived behavioral control (Ajzen, 1991). Thus, even if a stakeholder has formed strong attitudes towards certain actions, other contextual factors may still prevent the behavior.

## Figure 3.

Elaboration Likelihood Model supplementing Ajzen's (1991) Theory of Planned Behavior.



The ELM, therefore, supplements the Theory of Planned Behavior by providing guidance for one component of the theory (see Figure 3). Central route processing may not be sufficient on its own to produce evaluation influence at the behavioral level but may provide an important mediational contribution.

In summary, these findings highlight the importance of attitude strength towards supporting evaluation influence and highlight the need to consider these factors when designing evaluation products. While there is initial evidence that different data visualization designs may lead to different levels of persuasion, further work is needed to understand how and under what circumstances participants process these visualizations differently. Pandey et al. (2014) found that two different approaches to visualizing data – tables and graphical displays – resulted in different levels of persuasion among participants. However, which display was more effective for whom also depended on additional factors, such as initial beliefs. Those with strong attitudes counter to the presented conclusions tended to be more persuaded by data displayed in a table format while those without strong attitudes tended to find graphs more persuasive. Additionally, it is unclear to what extent participants in Pandey et al.'s (2014) study engaged in elaboration and central route processing since the study did not include both strong and weak evidence. If participants had been engaged in elaboration, they would have differentiated between strong and weak evidence. Therefore, while Pandey et al.'s (2014) findings provide an important first step towards demonstrating the importance of considering data visualization design for attitude change, further research is needed to understand how such visualizations can best support central route processing and attitude strength.

The final implication of the basic overview of the ELM may provide some initial guidance on factors that support central route processing. Petty and Cacioppo (1986) were able to categorize factors promoting central route processing into two overarching influences: motivation to elaborate and ability to elaborate. As such, evaluation products which incorporate features that promote motivation and ability to elaborate may result in a higher likelihood of central route processing.

Much of the research on the ELM has focused on factors that affect an individual's motivation and ability to elaborate, and thereby indirectly affect the extent of elaboration. While a wide range of factors have been examined, personal relevance, involvement, and need for cognition have been the most researched factors that increase motivation (e.g., Petty & Cacioppo, 1986; Pandey et al., 2014). Two of these factors, personal relevance and involvement,

are particularly applicable to evaluation given the role of stakeholders. Personal relevance is defined as the extent to which an individual is likely to be directly affected by an issue (Petty & Cacioppo, 1986). While a similar concept, individual involvement, refers to the personal commitment a viewer may have to the issue at hand – such as whether they will buy an advertised product (Wang, Wang, & Farn, 2009).

Both of these factors are likely to be relevant to evaluation stakeholders, who are "a group who has a 'stake' or vested interest in the process and outcome of the evaluation" (Preskill & Catsambas, 2006, p. 53). As such, they are likely to either be personally affected by issues explored in the evaluation or may have a personal commitment to the results of the study. For example, program staff may be personally affected by having to change how they go about doing their work based on the results of an implementation evaluation. Alternatively, program funders may continue funding based in part on the results of a summative evaluation. Additionally, participatory evaluation approaches incorporate practices that involve stakeholders in the evaluation process, thereby creating or heightening individual involvement. In this way, evaluation naturally incorporates these two motivational factors which should promote central route processing.

While Petty and Cacioppo (1986) provide a review of their substantial work looking into the effects of personal relevance and individual involvement on elaboration, more recent studies have also been conducted to replicate findings. For example, Langille et al. (2011) tested whether providing general versus personalized feedback as part of a workplace wellness program would affect attitudes towards physical activity. Findings suggest that while general feedback was more appreciated by those who were less fit, increasing the personal relevance of the messages was more effective at changing attitudes (Langille et al., 2011).

Similarly, Wang, Wang, and Farn (2009) measured individual involvement with a product in an advertising viewing context. The study repeatedly presented an advertisement on the same product but varied it either substantively (argument-relevant) or cosmetically (peripheral cue). Findings indicated that individuals with high product involvement tended to be more responsive to repeated advertisements with substantive changes while individuals with low product involvement tended to be swayed by repeated advertisements with cosmetic changes. Interpreted using the ELM, those with high product involvement tended to elaborate more and the substantive changes provided them with more information to consider (I like the product because I have substantial information indicating it is a good product). On the other hand, those with low product involvement tended to not elaborate and instead were more attune to the cosmetic changes (I like the product because the advertisement is aesthetically pleasing). This suggests that evaluation products or practices that are shaped to intentionally highlight personal relevance and increase individual involvement may be more likely to result in elaboration and central route processing.

Certain data visualization methods may be able to support heightening personal relevance and individual involvement of stakeholders in evaluation products. Byrne, Angus, and Wiles (2016) explored the role of socially constructed meanings in figurative elements of embellished graphs. Figurative elements were defined as "illustrations, photographs, cartoons, and schematic diagrams, where the meaning is based on the similarity of the shape of the representation to the shape of an external object or concept" (Bryne, Angus, & Wiles, 2016, p. 511). In a review of winning submissions to the Kantar "Information is Beautiful" 2014 Awards, Bryne et al. (2016) found that figurative elements were effective tools that attracted attention, oriented audiences to unfamiliar subjects, and made visualization memorable by embedding recognizable and familiar

symbols as anchors. Such an approach could be particularly powerful in an evaluation context by embedding culturally and organizationally relevant icons and symbols into visualizations. Such elements may be effective to signal the personal relevance of evaluation products and leverage pre-existing knowledge to provide scaffolding upon which to introduce new information gained from the evaluation. However, as Bryne and colleagues note, for such an approach to be effective it is critical to deeply understand the meaning such icons and images hold for the intended audience.

Alternatively, interactive graphs may be an effective method for increasing the individual involvement of stakeholders by turning the audience from "a passive consumer to an active participant" (Weissgerber et al., 2017, p. 20592). By interacting directly with data, the stakeholders can actively contribute to the evaluation analysis and findings by leveraging their own personal knowledge and curiosity, thereby "owning" a part of the evaluation. Additionally, interactive graphing may also be a method to increase personal relevance of the information by allowing the stakeholder to filter results to target the data most relevant to their needs.

Insights from these studies may also provide guidance around which data to visualize under what circumstances. For example, if personal relevance and individual involvement are salient to stakeholders, it may be more beneficial to visualize a variety of data so that stakeholders see a variety of arguments supporting the evaluative conclusions and recommendations. Similar to the substantially changed advertisements in Wang, Wang, and Farn's (2009) study, visualizing different data that support the same conclusion could provide stakeholders with additional evidence which they are likely to process through the central route. However, when personal relevance and individual involvement are unlikely to be salient to stakeholders, it may be more beneficial to focus on the aesthetic display of the information, and

potentially minimize the number of data graphed. However, research is needed to test these hypothetical relationships between visualization methods and motivational factors.

Finally, need for cognition, or the intrinsic tendency to engage in effortful thinking, has been the most examined personality trait in relation to increasing motivation to elaboration (Schumann, Kotowski, Ahn, & Haugtvedt, 2012). Cacioppo, Petty, and Morris' (1983; as cited in Petty & Cacioppo, 1986) original studies indicated that those high in need for cognition tend to be more intrinsically motivated to elaborate, and therefore more likely to take the central route. Need for cognition is a personal disposition towards engaging in effortful cognitive tasks. Those who have high need for cognition tend to enjoy such tasks and are pre-disposed to engage in them whereas those low in need for cognition tend to dislike such work and are pre-disposed to avoid expending cognitive effort. For example, one of their studies compared the argument appraisals of individuals who were high and low in need for cognition. Those who were high in need for cognition demonstrated greater scrutiny of arguments and more differentiation between strong and weak arguments than those low in need for cognition.

More recently, Kao (2011) replicated these findings as part of a larger study on message sidedness in advertising. Findings indicated that those high in need for cognition were more persuaded by two-sided persuasive messages (more arguments) than one-sided messages (fewer arguments), whereas the opposite was true for those low in need for cognition. While evaluators may not be able to affect the level of need for cognition among their stakeholders, gaining an understanding of this personality trait may give them a better understanding into the audience of their evaluation products. This in turn can provide design guidance, such as whether to provide an overview of all data collected or to focus on a few key findings.

In addition to increasing motivation, increasing ability to elaborate is also a viable way to enable central route processing of persuasive messages. One way to increase ability is through message repetition (Petty & Cacioppo, 1986). This approach is frequently used by advertisers, as seen in the Wang, Wang, and Farn (2009) study. In a review of ELM applications to consumer research, Lien (2001) notes that message repetition is "one of the most important variables influencing a person's ability to process issue-relevant arguments... [because it] provides more opportunities for argument scrutiny" (Lien, 2001, p. 302). Repetition can also provide additional opportunities for different types of elaboration. Malaviya (2007) showed that repeating the same ad allowed it to be placed in a different context, which in turn prompted two different types of elaboration. The ad itself focused on the attributes of the product, and repeated exposure gave additional opportunity to consider those attributes, a process known as item-specific elaboration. However, the context in which the ad was placed – meaning the ads surrounding the target ad – changed. This afforded additional opportunity for relational elaboration, or the consideration of the product in comparison to other products surrounding it. Thus, not only does repetition provide additional opportunity to engage in elaboration, it also enables a diversity of elaboration types.

The degree to which repetition may be incorporated into evaluation practice is dependent on the context of the evaluation. In some cases, evaluators can incorporate several evaluation products and processes, including reports, presentations, and stakeholder meetings, that can serve as venues for repeating evaluation findings. Other cases may be limited to a single evaluation report. However, there is potential even within a single evaluation report to repeat key messages throughout the document. However, as Malaviya's (2007) research indicates, mere repetition of the conclusions may not be sufficient. Instead, strategically varying the surrounding information

of the finding may support a diversity of elaboration types by providing additional context that may lead to new insights which further reinforce the importance of a finding. Additionally, findings from Malaviya's research underscore the importance of holistically examining how information is presented in evaluation products. Graphs are rarely standalone products, and the text and images surrounding them may color how stakeholders interpret and process the data these graphs represent. Such considerations may be important to preventing misinterpretations and over-generalizations.

Some factors detract, rather than promote, ability to elaborate. Distraction is one such factor. Petty, Wells, and Brock (1976, as cited in Petty & Cacioppo, 1986) conducted a series of studies that involved distraction tasks, such as counting clicks, for half of the participants. For these participants, the strength of their attitudes did not match the quality of arguments presented. For example, distracted individuals who were presented with weak arguments in favor of a tuition raise tended to express far more positive views of the issue than those who were not distracted. Alternatively, distracted individuals who were presented with strong arguments in favor of a tuition raise tended to be less favorable of the motion than those who were not distracted. Petty and Cacioppo (1986) argued that these findings were the result of an interrupted elaboration process that did not permit individuals to fully process and assess the arguments presented. Kohyama and Fujihara (1992) found support for this hypothesis by asking participants to both rate the extent to which they expended cognitive effort in considering the persuasive message and to list everything they were thinking during their reading of the message. Those who were distracted while reading the message demonstrated significantly lower levels of cognitive effort. These findings suggest that minimizing distraction can promote elaboration and central route processing of persuasive messages.

Stakeholders tend to have multiple demands on their time, and evaluators may be limited in the degree to which they can minimize distractions. Some practices, such as dedicating time to examining findings through scheduled meetings or data retreats, may be partially able to address such factors but are not always possible. However, this only serves to underscore the importance of designing evaluation products that account for these difficulties. In terms of data visualization, different graphing conventions may be leveraged to provide support in a variety of ways. For example, Bateman et al. (2010) found that embellished graphs were more attention grabbing and engaging. Strategic use of such conventions could help stakeholders focus in on key data and findings. On the other hand, minimalist approaches to visualization have been found to increase speed of interpretation (e.g., Cleveland & McGill, 1985; Gillan & Richman, 1994). Strategically using such graphs to quickly provide context or additional data that supplement the key data may present evaluation findings in a manner that most efficiently uses stakeholder's time. Rather than being opposing, these different visualization conventions might be most effective when used simultaneously. Bryne, Angus, and Wiles (2016) noted such mixed use of graphical conventions in some of the winning submissions in Kantar's "Information is Beautiful" 2014 awards.

Thus, motivation and ability both play a role in determining where on the elaboration likelihood continuum an individual lands. Importantly, insights from the ELM reframe the debate between minimalist and embellished graphs as two sides of the same coin. While embellished graphs may provide advantages in terms of increasing motivation to elaborate through making personal relevance more salient, minimalist graphs may increase ability to elaborate by quickly communicating data. Strategic use of each graphing convention may be guided by which factor – motivation or ability – is most likely to be in short supply in the particular context of an evaluation.

Kohyama and Fjihara (1992) began to explore the relationship between motivation and ability to elaborate by measuring motivation (operationalized as need for cognition) and ability (operationalized as distraction). Their initial results suggested that motivation significantly affected ability, suggesting that it plays a greater role in determination of elaboration. However, additional research is needed to disentangle this relationship, and to determine how and in what circumstances motivation and ability interact. If different graphing conventions are shown to support these two factors differentially, their combined use may offer an additional opportunity to explore such interactions.

Overall, these three initial insights provide guidance as to both situational and personal factors that may be important to consider when designing evaluation products intended to promote evaluation influence. The ELM provides guidance for how to promote attitude change that is resilient, persistent, and most likely to affect behaviors, factors which may play a role in the formation of evaluation influence pathways. Increasing the likelihood of central route processing is key, and evaluation products that enable elaboration are most likely to achieve this. It suggests that evaluators should be attuned to certain stakeholder attributes, such as the degree of involvement in the evaluation, extent to which it is personally relevant, and stakeholders' level of need for cognition. Additionally, strategies to efficiently use repetition and minimize distraction may be critical to helping stakeholders deeply process evaluative information. Such insights inform a variety of promising practices, from participatory approaches to specific data visualization methods.

#### **Other Roles for Evaluation in the Elaboration Context**

The ELM also includes more nuanced insights that may provide helpful guidance across evaluation's varied contexts. In addition to contributing to the extent of elaboration by

increasing motivation or ability to elaborate, a factor – in this case the evaluation products or processes – can play any of three other roles in a persuasive communication: a persuasive argument, a peripheral cue, or a biasing factor for elaboration. Which of the four roles the evaluation or its products may serve is determined by the pre-existing levels of motivation and ability to elaborate at the time of the communication, collectively known as the elaboration likelihood context (Petty & Cacioppo, 1986).

Evaluation is likely to serve as a persuasive argument when ability and motivation to elaborate are already high, indicating an elevated likelihood of automatic central route processing. This is the typical role that evaluation is seen to serve – providing information upon which to base attitudes and decisions. However, considering evaluation as a persuasive argument also highlights the need for such processes to produce strong evidence. While the central route has better outcomes (attitude resilience, persistence, and behavior intention), it also means greater differentiation between strong and weak arguments (Petty & Cacioppo, 1986). This has implications for both the rigor and credibility of evaluation studies. Studies that provide weak evidence or lack credibility are likely to be dismissed during elaboration, whereas strong and highly credible evidence is likely to sway audiences which engage in elaboration. Additionally, it is also important to remember that the very presence of an evaluation itself can serve as a persuasive argument if it represents an organization's genuine willingness to consider and improve its impact. This alone may add to a favorable assessment of the evaluand.

When motivation and ability to elaborate are low, evaluation is likely to take on the role of a peripheral cue via the peripheral route. Evaluation as a peripheral cue has already been acknowledged in the field as symbolic use of evaluation. However, an important implication of the ELM is that the peripheral route can still substantially change attitudes, though these changes

are not likely to be lasting. There are two possibilities that arise from this implication. First, stakeholders who process evaluations through the peripheral route may over- or under-exaggerate the evaluative conclusions because they do not actually process the supporting evidence for these conclusions. The direction of exaggeration will likely depend on whether the study's conclusions align with the stakeholder's previous assessment of the program. As such, if evaluators become aware that stakeholders have low motivation or ability to elaborate on the findings of an evaluation, they may need to take extra measures to ensure that stakeholder's perceptions are aligned with the study's conclusions.

On the other hand, evaluations processed through the peripheral route could open shortterm opportunities even in circumstances where the evidence is weak, which may serve as a steppingstone to further evaluations that might generate more rigorous or convincing data. However, evaluation as a peripheral cue also has ethical implications for evaluators. Even though research on the ELM – particularly in marketing research - has found many uses of peripheral cues, the Program Evaluation Standards clearly delineate propriety standards which include those related to human rights and respect (Yarbough, Shulha, Hopson, & Caruthers, 2011). It would be unethical to attempt to hide weak evidence using peripheral cues. Thus, strong caution is advised to evaluators seeking to apply the ELM in ways that may be inappropriate.

Finally, evaluation may serve as a biasing factor in the elaboration process. Biased elaboration occurs when processing of the information in a persuasive message occurs but is onesided such as only considering supporting arguments or only considering counterarguments rather than both (Petty & Cacioppo, 1986). Certain evaluation contexts may be likely to trigger such processing. For example, one phenomenon that can occur in evaluation is excessive evaluation anxiety, which refers to disproportionate and extreme anxiety that is induced in the

context of program evaluation (Donaldson, Gooler, & Scriven, 2002). Bechar and Mero-Jaffe (2014) note excessive evaluation anxiety can occur due to situational factors such as the threat of reducing funding, or due to dispositional factors such as previous negative experiences with evaluation. If under such circumstances elaboration does occur, it is possible that this processing will be one-sided due to the level of anxiety. Work by Ditto and Lopez (1992) indicates that fear appeals in persuasive messaging often results in "defensive processing", which dismisses frightening aspects of such messaging as unlikely to occur to them. It is possible, therefore, that under conditions of excessive evaluation anxiety, stakeholders may similarly minimize the implications of negative findings. Additionally, such processing may underlie several symptoms of excessive evaluation anxiety identified by Donaldson, Gooler, and Scriven (2002), such as withdrawal, resistance, and anger.

However, other reactions are also possible under conditions of excessive evaluation anxiety. Petty and Cacioppo (1986) note that under conditions where individual identity and core values are threatened, it is conceivable that participants shut down elaboration processes and operate purely on a peripheral route. This reaction is in alignment with research on the Social-Judgement Theory of Persuasion, where high involvement in an issue is associated with automatic rejection of counter-attitudinal messages (Sherif & Hovland, 1961). As clarified by Johnson and Eagly (1989), these research studies employed value-relevant involvement as an indicator of personal relevance, meaning that the issue explored in a persuasive communication was linked to core values of the participant. For example, Sherif et al. (1973) examined how students identified as Indian nationals by both themselves and their friends reacted to negative statements about India compared to Indian students who were not involved in Indian affairs. The

extremely unfavorable or moderately unfavorable. Students identified as Indian Nationalists were found to reject any negative statement regardless of how unfavorable it was while less involved students tended to be more non-committal (neither accept nor reject the statements).

Research studies on the Elaboration Likelihood Model, on the other hand, tended to focus on outcome-relevant involvement, meaning that personal relevance was determined by whether the individual would be affected by the outcome of an issue (e.g., a college student would be affected by a tuition increase). Therefore, the process by which attitudes are formed depends in part on the type of involvement. Translated to an evaluation context, high value-relevant involvement may be reflected in excessive ego involvement with the program model, a dispositional factor likely to trigger excessive evaluation anxiety (Donaldson, Gooler, & Scriven, 2002). If excessive evaluation anxiety is triggered because the evaluation threatens core values and identities of stakeholders, then it is possible that processing of evaluation findings will be more in alignment with Social-Judgement Theory and stakeholders may outright reject any counter-attitudinal messages.

Biased elaboration can also occur due to other contextual factors that are likely to occur in evaluations, for example in situations where the stakeholders have high levels of content knowledge. Petty and Cacioppo (1986) found high levels of previous knowledge to be a powerful biasing factor for elaboration because these individuals could utilize their expertise to form counterarguments or additional supporting arguments. In such cases, the direction of bias was based on whether the new information was aligned with previous knowledge. Therefore, when stakeholders include content experts, evaluators may need to prepare for biased assimilation of data. This has different implications depending on the findings. If findings match conventions, then stakeholders are likely to perceive the strength of the findings as stronger than

they are in actuality. In these situations, it may be beneficial to stress limitations and encourage careful consideration of next steps. If findings are counter to previous knowledge, stakeholders are likely to generate counterarguments. It may be beneficial in these cases to provide detailed results and to incorporate analyses that can answer any foreseen counterargument. Such efforts can be particularly helpful if evidence is strong enough to counter the counterarguments because these then become additional arguments in favor of the evaluative conclusions. Thus, awareness of factors that can bias the elaboration process may enable evaluators to preemptively respond.

Data visualization may play a role in how individuals process evaluation data in conditions in which biased elaboration is likely. Pandey et al. (2014) found that two different approaches to visualizing data – tables and graphical displays – resulted in different levels of persuasion among participants, but these relationships were moderated by participants' initial beliefs. When participants held strong attitudes counter to the conclusions presented, tables were found to be more persuasive. However, Pandey and colleagues found that in the absence of strong attitudes, graphs were more persuasive. Further examination suggested that while both sets of visualizations contained all the same data, they facilitated different processing. Tables were able to facilitate closer examination of other aspects of data which addressed potential counterarguments while graphs were more effective at highlighting key findings and showing general trends. As such, considering the likelihood of biased elaboration and audience members' knowledge and prior beliefs may be an effective guide for visualization design.

Key, however, to the nuance and complexity of the ELM is that a single factor can serve any of the four roles (persuasive argument, peripheral cue, facilitator/detractor of elaboration, or biasing factor) depending on the elaboration context (Petty & Cacioppo, 1986). For example, in low elaboration contexts, credibility is more likely to function as a peripheral cue leading to

acceptance of conclusions without consideration of the strengths of the supporting evidence ("she's an expert so the conclusions must be true"). Alternatively, in high elaboration contexts credibility is likely to be a persuasive argument which is examined and assessed when considering whether to accept the conclusion ("the guidance of an expert may have highlighted issues that would have otherwise undermined the study"; Petty and Cacioppo, 1986).

Finally, credibility can serve as a biasing factor. Tormala, Briñol, and Petty (2007) conducted research that explored how source credibility in a persuasive message effected favorable perceptions of a laundry product. Participants were randomly assigned to either a high credibility (government consumer research company) or low credibility (company selling the product) source. Additionally, the source was revealed either prior to the presentation of the message or after the participants had a moment to elaborate on the communication. Results indicate that participants reported more favorable views of the product when informed of a highly credible source prior to receiving the message. However, when source was revealed after the message, participants reported more favorable views of the product when the source was less credible. This suggests that prior knowledge of source credibility can bias the direction of elaboration processes. Together, these findings highlight the critical influence that an evaluation's credibility can have on acceptance of the evaluation's findings. These insights also bring attention to how the elaboration context might moderate the relationship between an evaluation's credibility and stakeholder reactions. However, it should be noted that the effects of credibility may be limited to attitude valence, as Azzam and Whyte (2018) found weak links between stakeholder's perceptions of the credibility of evaluative feedback and their intention to use this information.

In sum, understanding the elaboration context may be critical to understanding the role that an evaluation may play in convincing stakeholders to align their attitudes with the study's findings. Which role the evaluation serves in turn may have consequences for stakeholder's reactions, and attitude strength in particular, which may be important factors to consider in evaluation influence pathways. Motivation and ability to elaborate are key to central route processing, and other dispositional and contextual factors, such as prior content knowledge, distractions, and timing provide nuanced considerations to understanding the evaluation context. Having an in-depth understanding of which factors are likely to produce which role may allow evaluators to intentionally design evaluation products that produce certain reactions and to preemptively address other reactions that may arise due to the context.

#### **Research Questions**

The purpose of this research is to explore whether data visualization design informed by the Elaboration-likelihood Model (ELM) results in greater elaboration and evaluation influence. The hypothesized pathway to evaluation influence is displayed in Figure 4. Specifically, the research explores how embellished graphs, interactive data presentations, and minimalist graphs differ in terms of speed and accuracy of interpretation, audience perceptions, and resulting attitude change, attitude strength, and behaviors. Together, the results of this research will provide insights into whether and how data visualization can be used as a tool to promote evaluation influence as outlined in Mark and Henry's (2004) model. This research seeks to answer the following research questions:

- 1. How does motivation to elaborate affect the experience of data visualization?
  - a. How does motivation to elaborate affect **duration examining data visualization**? Does this depend on visualization approach or the strength of the evidence?

- b. How does motivation to elaborate affect how **engaging** participants find a visualization to be? Does this further depend on visualization approach and strength of the evidence?
- c. Does motivation to elaborate play a role in participants' **satisfaction** with the data visualization approach?
- 2. How does motivation to elaborate effect key interpretation and attitudinal outcomes of the presentation of evaluation findings?
  - a. Does motivation to elaborate play a role in participants' level of interpretation
     accuracy? Does this further depend on visualization approach and evidence strength?
  - b. Does motivation to elaborate affect participants' perceptions of the data as trustworthy?
     Do the strength of the evaluation findings and data visualization approach play a role in perceptions of trustworthiness?
  - c. Does **recall of evaluation findings** differ based on motivation to elaborate? Do the strength of the evaluation findings or data visualization approach play a role?
  - d. Does attitude change towards the evaluand differ based on motivation to elaborate, evidence strength, and/or data visualization approach? Does more global attitude change towards the same types of programs?
- 3. Does elaborating on evaluation findings result in greater likelihood of pathways of evaluation influence?
  - a. Do **attitude persistence and resilience** depend on motivation to elaborate, evidence strength, and/or data visualization approach?
  - b. Does **behavioral intent** differ based on motivation to elaborate, evidence strength, and/or data visualization approach?

- c. Does **donation behavior** differ based on motivation to elaborate, evidence strength, and/or data visualization approach?
- 4. Do lessons learned translate to interactive data presentations?
  - a. Are interactive data presentations perceived as **more engaging** than static data visualizations? Does this depend on motivation to elaborate or evidence strength?
  - b. Are interactive data presentations associated with **greater attitude change** than static data visualizations? Does this depend on motivation to elaborate or evidence strength?
  - c. Do interactive data presentations result in greater attitude persistence, resilience,

# donation behavioral intent, and donation behavior?

# Figure 4.

*Hypothesized pathways from data visualization design to attitude change and donation behaviors.* 



## **Overview of Research Design**

Overall, the research used an experimental design to answer the research questions, with a total of three phases. An overview of each phase and how it informs the subsequent phase is presented in the sections below (See Figure 5).

## Figure 5.

Relationships between the three research phases.



The first phase explored differences in experiences of data visualizations as well as interpretation and attitudinal outcomes based on motivation to elaborate. In particular, the study explored whether more engaging graphs promote greater attitude change, whether this attitude change further depends on the strength of the evaluation findings, and whether this experience differs by levels of motivation to elaborate.

Results from Phase 1 were used to further refine the hypotheses for Phase 2, which explored whether differences in motivation to elaborate result in greater attitude strength, behavioral intent, and behavioral action. These three factors are aligned with additional mechanisms of Mark and Henry's (2004) Theory of Evaluation Influence (cognitive and affective mechanisms and behavioral mechanisms, respectively), and together represent an influence pathway.

Finally, the findings of Phase 1 and 2 together were used to design an interactive data presentation intended to increase participant involvement in the display of the findings, thereby increasing motivation to elaborate. This in turn is anticipated to result in greater chance of

attitude change and influence pathways in Phase 3. The following chapters detail the methods and results of each phase of the research.

# CHAPTER 2 PHASE I: EXPERIMENTAL EXPLORATION OF ELABORATION LIKELIHOOD ON PARTICIPANTS' EXPERIENCE AND OUTCOMES OF DATA VISUALIZATION

The first phase of this research explores how differing levels of motivation to elaborate affect participants' experience as well as interpretation and attitudinal outcomes of embellished and minimalist data visualizations. In particular, the study addresses the first research question by exploring differences in time spent examining the graphs, degree to which the participant found the graph engaging, and participants' satisfaction with the visualization. Presented below are hypotheses for the roles of motivation to elaborate, data visualization approach, and evidence strength in how participants experience the visualization.

- 1. How does motivation to elaborate affect the experience of data visualization?
  - a. *Hypothesis 1.1:* Those high in motivation to elaborate will spend more time looking at the data visualization compared to those low in motivation to elaborate. This trend will be more pronounced for those presented minimalist graphs compared to embellished graphs.
  - b. *Hypothesis 1.2*: Embellished data visualizations will be perceived as more engaging compared to minimalist graphs. This trend will be more pronounced among those low in motivation to elaborate compared to those high in motivation to elaborate.
  - c. *Hypothesis 1.3*: Embellished graphs will be associated with higher satisfaction than minimalist graphs. This trend will be more pronounced for those low in motivation to elaborate.

Additionally, Phase 1 of the research explores key interpretation and attitudinal outcomes including accuracy of data interpretation, participants' perceptions of trustworthiness, the

memorability of the visualization and findings, and attitude change related to both the evaluand (specific attitude change) and that to the same types of programs more broadly (global attitude change). Presented below are hypothesized relationships between motivation to elaborate, data visualization approach, and evidence strength and interpretation and attitudinal outcomes.

- 2. How does motivation to elaborate effect key interpretation and attitudinal outcomes of the presentation of evaluation findings?
  - a. *Hypothesis 2.1:* Those high in motivation to elaborate will have higher accuracy of data interpretation compared to those low in motivation to elaborate. This trend will be less pronounced for embellished graphs compared to minimalist graphs.
  - b. *Hypothesis 2.2:* Those high in motivation to elaborate will perceive the data as trustworthy only when evidence is strong, regardless of visualization approach.
    Perceptions of data trustworthiness for those low in motivation to elaborate will depend on data visualization approach, such that minimalist graphs will be perceived as trustworthy regardless of evidence strength, while embellished graphs will be perceived as trustworthy only when presenting strong evidence.
  - c. *Hypothesis 2.3:* Those high in motivation to elaborate will have greater recall of the data compared to those low in motivation to elaborate. This trend will be more pronounced for those presented minimalist graphs compared to embellished graphs.
  - d. *Hypothesis* 2.4: Those high in motivation to elaborate will be more likely to elaborate on the data compared to those low in motivation to elaborate. This trend will be more pronounced for those presented minimalist graphs compared to embellished graphs.

Together, these questions address the hypothesized pathway to evaluation influence demonstrated in Figure 6.

Figure 6.



Hypothesized evaluation influence pathway explored in Phase 1.

**Design.** Phase 1 used a 2 (motivation to elaborate: low motivation vs. high motivation to elaborate) x 2 (data visualization approach: minimalist<sup>2</sup> vs. embellished<sup>3</sup> visualization) x 2 (evidence strength: weak evidence vs. strong evidence<sup>4</sup>) between-subjects experimental design. In an effort to more closely replicate conditions typical of the evaluation context, motivation to

<sup>&</sup>lt;sup>2</sup> Minimalist data visualization is defined as high data-ink data visualizations that eliminate "chart junk" (Tufte, 1983).

<sup>&</sup>lt;sup>3</sup> Embellished data visualization is defined as low data-ink data visualizations (Bateman et al., 2010).

<sup>&</sup>lt;sup>4</sup> Evidence strength refers to the difference between treatment and control conditions in the hypothetical evaluation findings.

elaborate was measured as the participants' natural pre-existing state rather than as a manipulated variable and was therefore not randomly assigned. Motivation to elaborate was measured as an individual characteristic and grouped as either "low" or "high" in elaboration likelihood.

Based on power analyses conducted using G\*Power, the total participant count needed for Phase 1 was 152 participants, with 19 participants in each of the 8 conditions to detect a medium effect size (f = 0.30). However, as Phase 1 included a delayed task to assess memorability which represented a significant opportunity for study attrition, the study oversampled to 216 participants, based on attrition rates reported in research by Christenson and Glick (2013).

## **METHODS**

**Recruitment.** The Phase 1 study leveraged the Amazon Mechanical Turk (MTurk) crowdsourcing site to recruit participants. Full recruitment text is included in Appendix B. The MTurk population is relatively similar to the general public, a common stakeholder group for evaluation (Difallah, Filatova, & Ipeirotis, 2018; Weiss, 1991). Difallah, Filatova, and Ipeirotis (2018) conducted a longitudinal study of MTurk workers to gain a deeper understanding of the MTurk population which accounts for the propensity to participate. Based on this research, MTurk workers tend to be slightly younger, slightly less likely to be married, and have slightly lower incomes compared to the US population (See Table 2, Difallah, Filatova, & Ipeirotis, 2018; U.S. Census Bureau, 2018).

Additionally, Heer and Bostock (2010) validated Amazon MTurk as a viable platform for the study of visualization design by replicating several laboratory visualization design studies. They found that MTurkers provide high quality responses and noted that using MTurk over

laboratory experiments had additional benefits of access to wider populations, cost reduction, and faster study completion. As a result, the Amazon MTurk platform represents an efficient approach to studying the effects of a variety of data visualization designs on elaboration and evaluation influence among a stakeholder group.

## Table 2.

MTurk demographics com	pared to US Population
------------------------	------------------------

<b>MTurk Workers</b>	<b>US Population</b>
20%	20%*
60%	40%*
80%	60%*
42%	52%
\$47,000	\$57,652
	MTurk Workers 20% 60% 80% 42% \$47,000

\*Statistic represents proportion of working age US population Source: Difallah, Filatova, & Ipeirotis, 2018; U.S. Census Bureau, 2018

**Procedure.** Overall, the study uses a simulation of a program evaluation report, with seven stages to the study procedure. First, participants were presented with a hypothetical scenario of a non-profit program called WISH that was geared towards improving the economic wellbeing and work-life balance of women through remote work. Then, participants took a short survey in order to be appropriately categorized as either high or low in motivation to elaborate and responded to a survey capturing their global attitudes towards programs that address economic wellbeing and work-life balance of women and their attitudes towards the hypothetical evaluand specifically. Third, participants were randomly presented one of four different hypothetical evaluation reports which differed in (1) the manner they presented the evaluation findings (embellished or minimalist graphs) and (2) the effect size of the evaluation findings (small effect size or large effect size). Fourth, participants responded to the attitude survey again. Fifth, participants responded to open-ended and survey questions regarding their interpretation of

the data, the degree to which they found the survey engaging, the trustworthiness of the evaluation, and their satisfaction with the way the findings were presented. Sixth, participants responded to demographic questions. Finally, after a week delay, participants were contacted again through email using the Amazon MTurk API and invited to participate in a follow-up questionnaire that checked whether they remembered the graph and data they had seen. Participants were then debriefed. Participants were paid up to \$3.40 in incentives through the Amazon MTurk portal, with \$2.40 upon completion of the initial survey and an additional \$1.00 upon completion of the follow-up survey.

#### Figure 7.

Seven stages of Phase 1 Study



*Scenario*. The hypothetical scenario presented to participants was developed in a manner that would naturally create two groups – one group for whom the program was relevant and would be highly likely to be motivated to elaborate and another for whom the program was not relevant and would not be likely to elaborate. Informed by early scenario pilots, the scenario achieved this by having the hypothetical program serve only women. Thus, the program should be significantly more relevant to female respondents then to male, which was confirmed through pilot testing using the motivation to elaborate questions outlined in the following section. The scenario is included in Appendix C.

Additionally, there is important context to consider regarding the remote work aspect of the scenario. Initial piloting of the scenario occurred in March 2020 while data for this research (Phases 1 - 3) was collected between October 2020 to February 2021. During this time, the

United States experienced the COVID-19 pandemic, which not only resulted in hundreds of thousands of deaths in the country, but also impacted the livelihoods of those who could not work remotely (Brynjolfsson et al., 2020). Given that household unemployment could impact participants' perceptions of the relevance of the program and their willingness to elaborate on evaluation findings, the scenario was piloted again in early December 2020 to ensure continued divides in relevance by gender.

Motivation to Elaborate. To understand participants' motivation to elaborate, level of personal relevance to the evaluand was assessed and was measured as the participants' natural pre-existing state rather than as a manipulated variable to more closely replicate the typical evaluation context. A significant body of research has shown that manipulating the level of personal relevance results in differing likelihood of elaboration (e.g., Langille et al., 2011; Petty & Cacioppo, 1986; Wang, Wang, and Farn, 2009). Two methods were used to determine level of personal relevance. First the study used Pandey et al. (2014) three-item measure of personal involvement, which explores: (i) the degree of interest in the topic, (ii) how much the individual feels the topic under discussion relates to her or his core values, and (iii) how much the individual feels the topic under discussion might have practical implications in her or his personal life. The measure used a seven-point semantic differential scale, and the three items were averaged. Items and internal consistency analyses are included in Appendix D. Based on survey results, participants were classified as low or high level of motivation to elaborate based on 40<sup>th</sup> and 60<sup>th</sup> percentile cut-off scores. Individuals whose scores fell between the 40<sup>th</sup> and 60<sup>th</sup> percentile were removed from the analyses.

Additionally, replicating previous research on the Elaboration Likelihood Model, participants were sampled by a personal characteristic that made the program more relevant. For

example, Petty and Cacioppo (1979, as cited in Petty & Cacioppo, 1986) manipulated personal relevance in their experiment by telling participants that a certain policy would be implemented in their school or a distant school. In alignment with this approach, the current study utilized gender as a secondary measure of motivation to elaborate as the hypothetical program was geared towards women. Thus, the program should have greater relevance to female participants.

*Attitudes Measure (Pre- and Post-).* Replicating an approach used by Mason and Azzam (2018), the attitudes towards the evaluand were measured through two scales: (1) a scale measuring global attitudes towards programs focused on improving the economic wellbeing and work-life balance of women using a seven-point semantic differential scale and (2) a scale measuring attitudes specifically towards the evaluand using a seven-point Likert scale (1 – strongly disagree to 7 – strongly agree). The scales were piloted with Amazon MTurk workers to ensure internal consistency (See Appendix E for scale items and scale internal consistency analyses).

*Evidence Strength.* Participants were randomly presented one of four presentations of the evaluation findings which differed by data visualization approach (minimalist graph or embellished graph; see section below) and by evidence strength (weak or strong evidence). Replicating Petty and Cacioppo's (1986) approach to determining strong and weak evidence, different effect sizes were piloted with Amazon MTurk workers by having pilot participants quantitatively rate the persuasiveness of each effect size in favor of the evaluand. Eight different effect sizes were piloted for both change in work-life balance and change in income. Based on average ratings of persuasion, and similarities in responses from both men and women pilot participants, weak evidence was defined as a 5% increase in work-life balance (from 10% to

15%) and an 8% income increase. Strong evidence was defined as a 45% increase in work-life balance (from 10% to 55%) and a 58% income increase.

*Data Visualization Approach.* The hypothetical evaluation findings were presented using two different approaches: (1) An embellished graph using iconography and (2) minimalist bar charts. The design of the embellished graph was informed by pilot testing and recommendations from a graphic designer. Additionally, like the Holmes graphs used in Bateman and colleagues' (2010) study, the embellished graphs were designed in a manner that reflected a positive value judgement of the evaluand (i.e., the program is effective) regardless of the effect size presented. Bateman and colleagues argue that one of the advantages of embellished graphs is that they can provide transparency about biases and value messages. To ensure this, pilot testing participants were asked whether they felt the researchers were trying to communicate a message through the visual and an open-ended question asking participants to describe what the visual is about. The minimalist graphs used bar graphs to present the study findings and minimized the use of non-data ink. Both graphs used the same titles and any other accompanying narratives (sub-titles, legends, interpretation text, etc.). The visualizations are included in Appendix F.

*Interpretation Survey*. Both qualitative and quantitative approaches were used to assess the accuracy of participants' data interpretation. In order to ensure that memory was not a confounding factor, participants were given the option to review the visual while completing the interpretation survey section of the study. First, an open-ended question asked participants to describe the study findings in their own words. Following this, participants completed a brief, multiple choice questionnaire to test the accuracy of their interpretation of the findings. Items asked questions about both specific data points and general trends. Given the need to match to specific data points, separate surveys were created for the weak evidence and strong evidence

conditions. The questions remained the same between the two conditions, but the response options changed to match the condition. Interpretation survey items for both conditions are included in Appendix G.

*Perceptions Survey.* Additionally, participants' perceptions of the trustworthiness of the data, the degree to which they found the visualization engaging, and overall satisfaction with the visualization were measured. Previous literature has suggested that both embellished and interactive graphs provide greater transparency than minimalist visualization conventions (Bateman et al., 2010; Weissgerber et al., 2016). Bateman and colleagues (2010) found that participants saw value messages in embellished charts significantly more often than in minimalist charts and argued that this increased transparency about underlying biases. However, it is possible that such value messages could also hurt the credibility of and trust in the evaluation study and findings. To measure this, participants were asked to rate the extent to which they trust (1) the evaluation findings and (2) find the evaluation study credible on a seven-point Likert scale (1 – not at all to 7 – completely).

Additionally, a 22-item User Engagement Questionnaire (VisEngage) developed by Hung and Parsons (2017) was used to assess the degree to which participants found the visual engaging. The questionnaire uses a seven-point Likert scale ranging from strongly agree to strongly disagree, and assesses 11 domains of engagement: aesthetics, captivation, challenge, control, discovery, exploration, creativity, attention, interest, novelty, and autotelism. VisEngage was created to assess engagement with interactive visualizations. As such, three items were removed, and other items slightly modified in order to remove mention of "interactivity" to better apply to the visualizations in this phase of research. Analyses showed high internal consistency across the final 19-items (see Appendix H). Finally, participants were asked to rate

their overall satisfaction with the visualization on a seven-point Likert scale (1 - strongly) disagree to 7 – strongly agree). Participants were given the option to review the visualization while completing the perceptions survey section of the study. All perceptions measures are included in Appendix H.

*Demographics.* In addition to questions such as age, gender, and educational attainment, the demographic section of the survey included questions regarding use of data visualization in everyday life. Participants were asked to indicate how frequently they see visualized data, create data visualizations, and use visualized data to make decisions (See Appendix I for all demographic questions).

*Time examining visual.* Previous research has argued that chart junk tends to decrease the speed of interpretation (e.g., Gillan & Richman, 1994; Schonlau & Peters, 2008; Seigrist, 1996). Therefore, in addition to the direct data collection, the study utilized survey metadata to determine the total duration of time participants spent looking at the visualization initially and in subsequent sections of the survey where they had the opportunity to review the visualization again (see interpretation and perceptions surveys above).

*Memorability Survey.* To assess the memorability of the visualization, participants were re-contacted after a week delay by sending an email using Amazon MTurk's API and Python programming (see Appendix B for full email text). To ensure only participants who had completed the first survey were eligible to complete the follow-up, each participant was assigned a worker qualification in the Amazon MTurk portal indicating their previous participation in the study. The study oversampled to account for attrition rates, and attrition across conditions was analyzed to ensure equivalent groups. Participants were presented with the description of the evaluand again and reminded that they had seen the findings of an evaluation study about the

program a week before. They were asked an open-ended question to describe the study findings in their own words. Then, participants were presented with each of the four data visualizations and asked to indicate whether they had seen the visualization previously.

**Analysis Approach.** Two approaches were used to analyze the study data. First, two three-way MANOVA analyses were conducted utilizing IBM Statistics SPSS Version 21 to explore the relationship between the three independent variables (motivation to elaborate level, data visualization approach, and evidence strength) as well as their interactions with the dependent variables. MANOVA analyses were utilized to minimize alpha inflation and dependent variables were conceptually grouped into variables related to experience of the study materials (duration examining materials, engagement with materials, and satisfaction with visual) and outcome variables (perceptions of trustworthiness, accuracy of interpretation, global and specific attitude change, and memory of findings). Second, participants' qualitative responses describing the evaluation findings both immediately after seeing the study materials and after a week delay were coded for accuracy of interpretation as well as emergent themes. These qualitative codes were then assigned numeric nominal values and the relationship between qualitative themes and the independent variables (motivation to elaborate, data visualization approach, and evidence strength) were examined through chi-square tests of independence. Results of the chi-square analyses were then compared to accuracy of interpretation and memory of findings MANOVA results in a convergent parallel mixed method approach.

#### RESULTS

**Population.** The study recruited 216 participants from Amazon Mechanical Turk (MTurk), a crowdsourcing platform. The study participants were limited to Amazon MTurk workers who reside in the United States to minimize potential confounding factors and the risk
of language barriers as the study was conducted in American English. A total of 14 participants were removed from the sample because their responses to the gender demographic question did not match the gender reported in their Amazon Mturk profile, reducing the sample size to 202. Additionally, a follow-up survey was completed after one week, with 114 participants responding for a 56.4% retention rate. There were no significant differences in retention between conditions ( $\chi^2(3) = 1.072$ , p = 0.784).

Based on responses to demographic questions, half of the initial sample were female. However, males were slightly more likely to respond to the one-week follow-up survey, making up 53.5% of the 114 follow-up respondents. Retention did not differ significantly by gender  $(\chi^2(1) = 1.156, p = 0.282).$ 

Most respondents in the initial sample were either 25-34 years old (38.6%) or 35-44 years old (26.7%, See Table 3). No participant reported being over 74 years old. Based on Pearson Chi-Square analyses, there were no significant differences in age group based on study condition in the initial sample ( $\chi^2(18) = 14.216$ , p = 0.715) nor in the follow-up sample ( $\chi^2(18) = 15.137$ , p = 0.653).

#### Table 3.

Phase I	l participants l	by age groups
---------	------------------	---------------

	Initial Sample	Follow-up Sample
Age Group	( <i>N</i> = <b>202</b> )	( <i>N</i> = 114)
18-24 years old	4.0%	3.5%
25-34 years old	38.6%	33.3%
35-44 years old	26.7%	28.1%
45 – 54 years old	17.3%	16.7%
55-64 years old	10.4%	13.2%
65 – 74 years old	2.5%	4.4%

All participants in the initial sample reported having earned at least a high school degree; most (44.1%) had completed a bachelor's degree and an additional 23.3% had completed a master's degree or professional degree (See Table 4). Similar trends held for the follow-up sample. Based on Pearson Chi-Squared analyses, there were no significant differences in education level across conditions in the initial sample ( $\chi^2(9) = 8.036$ , p = 0.530) nor in the follow-up sample ( $\chi^2(9) = 8.175$ , p = 0.517).

### Table 4.

	Initial Sample	Follow-up
Education	( <i>N</i> = <b>202</b> )	<b>Sample</b> ( <i>N</i> = 114)
High School Diploma/ GED	13.4%	15.0%
Associate's Degree/ Some College	19.3%	16.8%
Bachelor's Degree	44.1%	46.9%
Master's Degree/ Professional Degree	23.3%	21.2%

Participants' highest level of competed education.

Demographic questions also explored participants' familiarity with data visualization through three questions. Overall, results indicate that participants tended to see visualized data (M = 3.99, SD = 1.16) and use data displayed in charts to make decisions (M = 3.51, SD = 1.26)on a weekly basis but did not typically create data visualizations themselves (M = 2.83, SD =1.49, See Table 5). There were no significant differences in mean responses to these questions between those who completed the follow-up survey and those who did not (t(199) = -1.245, p =0.215; t(199) = 0.503, p = 0.222; t(199) = 1.275, p = 0.204, respectively).

 Table 5.

 Initial sample participants' experience with data visualization (N=201).

Question	Never	Less than once per month	Once a month	Once a week	Once a day	More than once a day
In a typical month, how often do you see data displayed in a chart?	4.0%	5.0%	17.8%	45.0%	17.8%	10.4%
In a typical month, how often do you use data displayed in a chart to make decisions?	5.4%	17.8%	21.8%	36.6%	11.4%	6.9%
In a typical month, how often do you create a chart to display data?	25.2%	21.3%	17.8%	20.8%	10.9%	4.0%

In order to determine if differences in participants' data visualization experience differed by gender, an average score was calculated using the responses to the three items. Independent samples t-test analyses determined there were no statistically significant differences in data visualization experience based on gender (*Levene's* F = 3.428, p = 0.066, t(200) = 1.661, p =0.098). Similarly, a one-way ANOVA analysis revealed no statistically significant differences in data visualization experience based on age (F(6, 195) = 0.789, p = 0.580).

Despite random assignment, one-way ANOVA analyses revealed statistically significant differences in data visualization experience between study conditions (F(3, 198) = 3.930, p = 0.009). Post-hoc Bonferroni comparisons revealed that those who were assigned to the minimalist visualization displaying weak evidence had significantly higher experience with data visualization than those who were assigned to view embellished visualization displaying strong evidence (Mean difference = 0.716 [95% CI, 0.149 to 1.283], p = 0.006).

*Motivation to Elaborate*. Motivation to elaborate was calculated by averaging the responses of the three seven-point semantic differential scale items (M = 5.16, SD = 1.38, N = 202). Independent samples t-test revealed that motivation to elaborate significantly differed by gender (*Levene's* F = 7.169, p = 0.008, t(193.404) = -5.505, p < 0.001), such that female participants reported higher levels of motivation to elaborate than male respondents (see Table 6). Similar trends held for the follow-up sample (*Levene's* F = 6.048, p = 0.015, t(110.355) = -4.624, p < 0.001). Given the alignment of the two measures of motivation to elaborate, further analyses were conducted utilizing the mean score of the three-item motivation to elaborate questionnaire.

#### Table 6.

Motivation to elaborate scores by participant gender.

	1	Initial Samp	le	Follow-up Sample		
Gender	N	Mean	SD	N	Mean	SD
Female	101	5.66	1.17	53	5.71	1.18
Male	101	4.68	1.41	60	4.55	1.45

Average motivation to elaborate scores did not differ significantly based on condition (F(3, 194) = 0.081, p = 0.970). However, one-way ANOVA analyses revealed that while motivation to elaborate did not differ by age (F(5, 195) = 1.930, p = 0.091), there were differences based on education (F(3, 197) = 4.788, p = 0.003). Post hoc Bonferroni comparisons showed that those with a master's degree reported significantly higher levels of motivation to elaborate than those with a high school diploma/GED (Mean difference = 0.928 [95% CI, 0.081 to 1.776], p = 0.023), an associate's degree or some college (Mean difference = 0.951 [95% CI, 0.0212

to 1.289], p = 0.039). The trend held for the follow-up sample as well (F(3, 109) = 4.170, p = 0.008).

In order to differentiate individuals with low and high levels of motivation to elaborate, cut-off points were used. Those with an average motivation to elaborate score at or below the  $40^{\text{th}}$  percentile (score of 5.067) were designated as having low motivation to elaborate while those at or above the  $60^{\text{th}}$  percentile (score of 5.667) were designated as having high motivation to elaborate. Based on this, 81 individuals had low motivation to elaborate while 121 individuals had high motivation to elaborate levels. Based on Pearson Chi-squared analyses, levels of motivation to elaborate did not differ across conditions ( $\chi^2(3) = 1.587, p = 0.662$ ).

Analysis of Data Visualization Experience. Two three-way MANOVA analyses were conducted in order to examine the effects of level of motivation to elaborate, data visualization approach and evidence strength on experience of the visualizations (duration examining visual, engagement with the visual, and satisfaction with the visual) and key outcomes (perceptions of trustworthiness, accuracy of interpretation, global and specific attitude change, and memory of findings).

The first three-way MANOVA utilized the three continuous dependent variables related to experience, which are summarized in Table 7. Distribution of participants across conditions after removal of outliers are included in Table 9. Independence of observations was attained through study design.

Table 7.					
Summary of dependent	t variables	related to	o data	visualization	experience

Dependent Variable	N	Mean	SD	Min	Max	Skew	Kurtosis
Duration	202	576.03	326.78	112.00	1997.00	1.43	2.45
Engagement	202	4.89	1.14	1.47	7.00	-0.54	-0.09
Satisfaction	202	5.61	1.29	1	7	-1.40	2.17

There was a linear relationship between the dependent variables, as assessed by scatterplot. However, Pearson's correlations indicated multicollinearity between the satisfaction and engagement variables for the low motivation to elaborate, weak strength embellished graph group (Pearson's r = 0.919). To address this, satisfaction scores were removed from further analyses. A total of 14 univariate outliers were identified through boxplot for values greater than 1.5 box-lengths from the edge of the box and were removed listwise. There were no multivariate outliers in the data, as assessed by Mahalanobis distance (critical value = 13.82, p > 0.001). Multivariate normality was assessed using Shapiro-Wilk test using a Bonferroni correction (p < 0.006). A total of 4 instances of broken multivariate normality assumption was detected (See Table 8). However, as the three-way MANOVA is robust to Type I error, analyses proceeded without transformation to the dependent variables (Bray & Maxwell, 1985; Weinfurt, 1995).

#### Table 8.

Instances of broken multivariate normality assumption assessed by Shaprio-Wilk statistic with Bonferroni correction.

	Condition					
Motivation				Shaprio-		
to	Visualization	Evidence	Dependent	Wilk		
Elaborate	Approach	Strength	Variable	Statistic	Df	Sig.
Low	Minimalist	Weak	Duration	0.791	21	< 0.001
High	Minimalist	Strong	Engagement	0.877	28	0.003
High	Embellished	Weak	Engagement	0.874	34	0.001
High	Embellished	Strong	Duration	0.698	33	< 0.001

The assumption of adequate sample size was met, with each group having a sample of 17 or greater (needed at least a sample size of 2 per group; See Table 9 for sample sizes). The homogeneity of covariance matrices assumption was broken, as assessed by Box's M test (Box's M = 59.98, F(21, 75440.91) = 2.746, p < 0.001). As a result, multivariate analyses were

determined utilizing Pillai's criterion as it is more robust to unequal covariance matrices (Olson, 1976). Homogeneity of variance was not equal based on Levene's Test (Duration: F(7, 179) = 4.83, p < 0.001; Engagement: F(7, 179) = 2.375, p = 0.024). As a result, the significance level was adjusted to 0.01.

No significant 3-way interaction was found between data visualization approach, level of motivation to elaborate, and evidence strength (Pillai's Trace = 0.033, F(2, 178) = 2.996, p = 0.053, partial  $\eta^2 = 0.033$ ).

#### Table 9.

	Motivation to	Strength	
Chart Type	Elaborate	Туре	N
	Low	Weak	19
Minimalist		Strong	21
winninansi	High	Weak	26
	Ingn	StrengthIaborateTypeWeakStrongWeakStrongWeakStrongWeakStrongStrongStrong	25
	Low	Weak	17
Fmbellished	Low	Strong	18
Linoemsneu	High	Weak	33
	111511	Strong	28

Final sample size per condition for data visualization experience MANOVA analyses.

Multivariate tests revealed a significant 2-way interaction between level of motivation to elaborate and evidence strength (Pillai's Trace = 0.060, F(2, 178) = 5.716, p = 0.004, partial  $\eta^2 = 0.060$ ; See Figure 8). Test of between-subjects effects indicated that the 2-way interaction was significant for duration examining materials, F(1, 179) = 10.879, p = 0.001, partial  $\eta^2 = 0.057$ , but not for engagement, F(1, 179) = 0.770, p = 0.381, partial  $\eta^2 = 0.004$ . Pairwise comparisons indicate that individuals who were highly motivated to elaborate spent 141.62 seconds (95% CI,

46.69 to 236.56) longer when examining materials with weak evidence compared to strong evidence, a statistically significant finding (p = 0.004). For those who were low in motivation to elaborate, there were no significant differences in duration examining materials with weak evidence compared to strong evidence [Mean difference = -108.55 seconds, CI 95%(-224.26 to 7.163), p = 0.066].

#### Figure 8.

Mean duration examining visual by level of motivation to elaborate and evidence strength.



Multivariate tests also revealed a significant main effect of motivation to elaborate (Pillai's Trace= 0.271 F(2, 178) = 33.161, p < 0.001, partial  $\eta^2 = 0.271$ ). Test of betweensubjects effects indicated a main effect of motivation to elaborate for engagement, F(1, 179) =66.611, p < 0.001, partial  $\eta^2 = 0.271$ , but not for duration examining materials, F(1, 179) =0.267, p = 0.606, partial  $\eta^2 = 0.001$ . Those who were highly motivated to elaborate reported significantly higher engagement scores [M = 5.38, CI 95% (5.20 to 5.56), SD = 0.092] than those who were low in motivation to elaborate [M = 4.19, CI 95%(3.97 to 4.41), SD = 0.113; See

Figure 9].

# Figure 9.

Mean engagement score by level of motivation to elaborate.



Given that satisfaction scores were removed from the three-way MANOVA due to multicollinearity with engagement scores, a separate three-way ANOVA analysis was completed to assess the relationship between level of motivation to elaborate, data visualization approach, and evidence strength and satisfaction with the visualization. Boxplots revealed there were 14 outliers assessed as a value greater than 3 box-lengths from the edge of the box. As these were neither data entry nor measurement errors, they were retained in the analysis. Shapiro-Wilk test of normality revealed that satisfaction scores were not normally distributed (p < 0.05) for all but one group (low motivation to elaborate, minimalist visualizations with weak evidence p = 0.177). As a result, a reflect and logarithmic transformation was applied to the satisfaction scores, with the resulting correction to skew summarized in Table 10. There was homogeneity of variances for transformed satisfaction scores as assessed by Levene's test for equality of variances, F(7,

193) = 0.721, p = 0.655.

## Table 10.

Comparison of skew and kurtosis between original and transformed satisfaction scores.

Condition		Ske	ewness	Kurtosis		
Visualization	Elaboration	Evidence	<u> </u>		o · · · 1	
	Motivation	Strength	Original	Transformed	Original	Transformed
Minimalist	Low	Weak	-0.383	-0.685	-0.024	-0.118
winningi	LOW	Weak	(0.501)	(0.501)	(0.972)	(0.972)
Minimalist	Low	Strong	-1.164	0.269	0.627	0.147
wiiiiiiiaiist	LOW	Suong	(0.501)	(0.501)	(0.972)	(0.972)
Minimalist	Iliah	Waal	-0.509	-0.081	-0.671	-1.394
Minimansi	High	weak	(0.456)	(0.456)	(0.887)	(0.887)
Minimalist	Iliah	Steena	-0.075	-0.610	-0.270	-0.634
Minimansi	nign	Strong	(0.441)	(0.441)	(0.858)	(0.858)
Emballished	Low	Week	-1.004	-0.122	0.147	-0.170
Embernsheu	LOW	weak	(0.536)	(0.536)	(1.038)	(1.038)
Emballished	Low	Steena	-1.597	0.211	2.979	-0.139
Embernsheu	LOW	Suong	(0.512)	(0.512)	(0.992)	(0.992)
Emballished	Iliah	Waal	-1.882	0.549	4.139	-0.202
Embernsheu	nigii	weak	(0.403)	(0.403)	(0.788)	(0.788)
Each alliab a d	II: ah	Ctuere e	-1.311	0.061	1.915	-0.312
Embellished	High	Strong	(0.409)	(0.409)	(0.798)	(0.798)

Test of between-subject effects revealed no significant three-way interaction between level of motivation to elaborate, data visualization approach, and evidence strength, F(1, 193) =2.332, p = 0.128. Similarly, no significant two-way interactions were found (p < 0.05). However, a statistically significant main effect of level of motivation to elaborate was identified, F(1, 193)= 21.124, p < 0.001. Comparing original satisfaction scores, those who were low in motivation to elaborate tended to have significantly lower satisfaction scores in comparison to those who were high in motivation to elaborate (M = 5.11, SD = 1.48 and M = 5.93, SD = 1.03, respectively; See Figure 10).

#### Figure 10.

Statistically significant differences in satisfaction with visual based on motivation to elaborate.



Analysis of Outcomes. The second three-way MANOVA utilized five continuous dependent variables related to key outcomes, summarized in Table 11. Given that memory was an outcome explored, the analysis focused on the 114 participants who completed the follow-up survey. There was a linear relationship between the dependent variables, as assessed by scatterplot and no evidence of multicollinearity, as assessed by Pearson correlation (|r| < 0.9). There were 16 univariate outliers identified through boxplot for values greater than 3 box-lengths from the edge of the box and three multivariate outliers in the data, as assessed by Mahalanobis distance (critical value = 20.52, p > 0.001). As these were neither data entry nor measurement errors, outliers were retained for the analyses.

Dependent Variable	N	Mean	SD	Min	Max	Skew	Kurtosis
Trustworthiness	114	5.62	1.16	1.00	7.00	-1.69	4.04
Accuracy of Interpretation	114	3.75	1.28	0.00	5.00	-0.85	-0.21
Global Attitude Change	114	0.16	0.53	-2.00	2.40	0.67	5.13
Specific Attitude Change	114	0.29	0.59	-0.80	3.20	1.96	3.58
Accuracy of Memory	114	-0.94	1.63	-4.00	3.00	0.286	-0.126
Accuracy of Memory 2	114	2.81	1.50	0.00	5.00	-0.226	-1.03

**Table 11.**Summary of Dependent Variables Related to Data Visualization Experience

Multivariate normality was assessed using Shapiro-Wilk test using a Bonferroni correction (p < 0.006). A total of 12 instances of broken multivariate normality assumption was detected (See Table 12). However, as the three-way MANOVA is robust to Type I error, analyses proceeded without transformation to the dependent variables (Bray & Maxwell, 1985; Weinfurt, 1995). Adequate sample size assumption was met, with each group having a sample of 10 or greater (needed at least a sample size of 5 per group; See Table 13 for sample sizes). Homogeneity of covariance matrices assumption was broken, as assessed by Box's M test (Box's M = 207.82, F(105, 9684.53) = 1.834, p < 0.001). As a result, multivariate analyses were determined utilizing Pillai's criterion as it is more robust to unequal covariance matrices (Olson, 1976). The assumption of homogeneity of variance was broken for trustworthiness [F(7, 106) =3.087, p = 0.005].

# Table 12.

Instances of broken multivariate normality assumption assessed by Shaprio-Wilk statistic with Bonferroni correction (p<0.006).

	condition					
Motivation	Visualization	Evidence	-	Shaprio-		
to Elaborate	Approach	Strength	Dependent Variable	Wilk	Df	Sig.
Low	Minimalist	Strong	Accuracy of Interpretation	0.623	12	< 0.001
Low	Embellished	Strong	Trustworthiness	0.732	10	0.002
Low	Embellished	Strong	Global Attitude Change	0.757	10	0.004
Low	Embellished	Strong	Specific Attitude Change	0.678	10	< 0.001
High	Minimalist	Weak	Specific Attitude Change	0.638	17	< 0.001
High	Minimalist	Strong	Trustworthiness	0.772	14	0.002
High	Minimalist	Strong	Accuracy of Interpretation	0.790	14	0.004
High	Minimalist	Strong	Global Attitude Change	0.773	14	0.002
High	Minimalist	Strong	Specific Attitude Change	0.774	14	0.002
High	Embellished	Weak	Accuracy of Interpretation	0.800	21	0.001
High	Embellished	Strong	Trustworthiness	0.819	18	0.003
High	Embellished	Strong	Global Attitude Change	0.706	18	< 0.001

# Condition

# Table 13.

Final sample size per condition for outcome MANOVA analyses.

	Motivation to	Strength	
Chart Type	Elaborate	Туре	N
	Low	Weak	11
Minimalist	Low	Strong	12
	High	Weak	17
		Strong	14
	Low	Weak	11
Embellished	Low	Strong	10
	High	Weak	21
	mgn	Strong	18

No significant 3-way interaction was found between data visualization approach, level of motivation to elaborate, and evidence strength (Pillai's Trace = 0.080, F(5, 102) = 1.770, p = 0.126, partial  $\eta^2 = 0.080$ ).

Multivariate tests indicated a significant 2-way interaction between level of motivation to elaborate and evidence strength, but not data visualization approach and evidence strength nor data visualization approach and level of motivation to elaborate (See Table 14).

## Table 14.

Results of multivariate tests for 2-way interactions.

Interaction	Pillai's			Partial
	Trace	<b>F-value</b>	Sig.	$\eta^2$
Motivation to elaborate x Evidence Strength	0.104	2.371	0.044	0.104
Data Visualization Approach x Evidence Strength	0.038	0.799	0.533	0.038
Data Visualization Approach x Motivation to Elaborate	0.030	0.630	0.677	0.030

Follow-up ANOVA between-subject effects revealed significant interaction effects of level of motivation to elaborate and evidence strength for trustworthiness (F(1, 106) = 6.893, p = 0.013, partial  $\eta^2 = 0.057$ ), but not for the remaining outcome variables (p < 0.05). Based on pairwise comparisons, those low in motivation to elaborate have a -1.01 [CI95%(-1.63 to -0.39), p < 0.002] point lower trustworthiness score when presented weak evidence compared to strong evidence, whereas those high in motivation to elaborate do not have a significant difference in their trustworthiness scores based on evidence strength [Mean difference = 0.002, 95% CI(-0.493 to 0.498), p = 0.993; See Figure 11].

#### Figure 11.



Mean trustworthiness scores based on level of motivation to elaborate and evidence strength.

Multivariate tests revealed a significant main effect of level of motivation to elaborate (Pillai's Trace = 0.189, F(5, 102) = 4.746, p = 0.001, partial  $\eta^2 = 0.189$ ) and evidence strength (Pillai's Trace = 0.162, F(5, 102) = 3.938, p = 0.003, partial  $\eta^2 = 0.162$ ), but not data visualization approach (Pillai's Trace = 0.023, F(5, 102) = 0.470, p = 0.798, partial  $\eta^2 = 0.023$ ). Follow-up tests of between-subjects effects revealed statistically significant differences in accuracy of interpretation, (F(1, 106) = 6.069, p = 0.049, partial  $\eta^2 = 0.036$ ) and specific attitude change, F(1, 106) = 1.692, p = 0.027, partial  $\eta^2 = 0.045$ ) based on level of motivation to elaborate. The remainder of the outcome variables did not differ significantly based on level of motivation to elaborate (p < 0.01).

In terms of accuracy of interpretation, pairwise comparisons indicate that those who were low in motivation to elaborate had significantly higher accuracy of interpretation than those who were high in motivation to elaborate (See Figure 12).

# Figure 12.

Differences in accuracy of interpretation scores by level of motivation to elaborate.



In terms of specific attitude change, pairwise comparisons indicate that those low in motivation to elaborate had significantly greater attitude change specific to the evaluand than those who were high in motivation to elaborate (See Figure 13).

# Figure 13.

Differences in attitude change specific to evaluand by level of motivation to elaborate.



Additionally, memorability was assessed through a recognition task where participants were asked to identify the visualization they had previously seen out of the four possible prompts (strong evidence with minimalist visualization, weak evidence with minimalist visualization, strong evidence with embellished visualization, weak evidence with embellished visualization).

Pearson's Chi-square analyses revealed significant differences in recognition of materials based on level of motivation to elaborate ( $X^2(3) = 8.308, p = 0.040$ , Cramer's V = 0.270). Those that were highly motivated to elaborate were significantly more likely to have no memory of the materials (not recognize any aspect of the materials) than expected, as assessed by examination of standardized adjusted residuals using small table cutoff of greater than 2 in absolute value (Agresti, 2007; see Table 15).

#### Table 15.

	Low Motivation to	High Motivation to
<b>Recognition Type</b>	Elaborate	Elaborate
No recognition	12	38
	(-2.8)	(2.8)
Recognize data visualization	10	12
approach only	(0.7)	(-0.7)
Recognize evidence strength	8	7
only	(1.3)	(-1.3)
Recognize correct chart (both	1.4	12
data visualization approach	14	13
and evidence strength)	(1.6)	(-1.6)

Crosstabulation of chart recognition and level of motivation to elaborate.

Note. Adjusted residuals appear in parentheses below observed frequencies.

Pearson Chi-squared analyses also revealed that recognition of the materials differed significantly based on the data visualization approach utilized ( $X^2(3) = 23.014$ , p < 0.001, Cramer's V = 0.449). Examination of standardized adjusted residuals revealed that those

who had been assigned to the embellished visualization condition recognized only the strength of the findings but not data visualization approach more frequently than expected (See Table 16; Agresti, 2007). Conversely, those who were assigned to the minimalist visualization condition recognized the correct visualization (both in visualization approach and evidence strength) more frequently than expected.

#### Table 16.

	Minimalist Data	<b>Embellished Data</b>
<b>Recognition Type</b>	Visualization Condition	Visualization Condition
No recognition	21	29
	(-1.0)	(1.0)
Recognize data visualization	13	9
approach only	(1.2)	(-1.2)
Recognize evidence strength	0	15
only	(-3.9)	(3.9)
Recognize correct chart (both	20	7
data visualization approach	20	1
and evidence strength)	(3.2)	(-3.2)

Crosstabulation of chart recognition and data visualization approach.

Note. Adjusted residuals appear in parentheses below observed frequencies.

Finally, Pearson Chi-square analyses revealed significant differences in recognition of materials based on the strength of evidence presented ( $X^2(3) = 20.373, p <$ 

0.001, Cramer's V = 0.423). Standardized adjusted residuals indicated that those who were in the weak evidence condition did not recognize the materials at all more frequently than expected while those who were in the strong evidence condition had accurately identified the correct materials (both evidence strength and visualization approach) more frequently than expected (See Table 17; Agresti, 2007).

#### Table 17.

<b>Recognition Type</b>	Weak Evidence condition	Strong Evidence Condition
No recognition	36	14
	(3.7)	(-3.7)
Recognize data visualization	12	10
approach only	(0.2)	(-0.2)
Recognize evidence strength	7	8
only	(-0.5)	(0.5)
Recognize correct chart (both	5	22
data visualization approach	5	22
and evidence strength)	(-4.1)	(4.1)

Crosstabulation of chart recognition and evidence strength condition.

*Note*. Adjusted residuals appear in parentheses below observed frequencies.

**Qualitative Analysis**. In addition to quantitative assessment of accuracy of interpretation, 167 qualitative responses to the prompt "Please describe the study findings in your own words" were coded for accuracy and emergent themes by the researcher. An additional 35 responses were removed as uninterpretable due to being one-word or off-topic.

Accuracy was explored in terms of accuracy of described trends and accuracy of provided statistics. In terms of trend accuracy, responses were coded as accurate if they describe the program outcomes as positive but did not utilize specific statistics. Table 18 below provides sample responses for each code. Accurate trend was the most common response type (n = 78, 46.7%). Conversely, the code 'inaccurate trends' described responses that indicated the outcomes of the program were negative, such as that WISH program participants did not increase their income or that fewer WISH participants experienced work-life balance after participating in the program; however, no responses were coded as including an inaccurate trend.

Responses which included statistics were coded depending on whether they accurately reported the statistics for their condition (weak or strong evidence). This could include either

reporting the exact statistics that were included in the description of the evaluation findings they read or statistics they calculated themselves as differences between those presented. For example, a participant in the weak evidence strength condition could accurately report that satisfaction in work-life balance increased from 10% to 15% or that it increased 5%. A total of 40 responses (24.0%) included accurate descriptions of statistics while 3 responses (1.8%) included inaccurate descriptions of statistics.

Responses were coded as partially correct if the content of the responses included both accurate and inaccurate descriptions, such as a mix of statistics, some of which were correct and some of which were incorrect. This also included responses which inaccurately identified the outcomes described in the condition, but correctly described a positive trend. A total of 20 responses (12.0%) were coded as partially correct.

A theme that emerged in the partially accurate responses was that participants extended the findings to be inclusive of additional or alternative outcomes that were not included in the description of findings they had been presented. These fell into two sub-codes: (1) extension of work-life balance and (2) increased productivity. The first sub-code was utilized when participant responses included alternative or additional outcomes in place of the concept of work-life balance. For example, several participants described the work-life balance finding as increased quality of life, increased satisfaction with work, or better moods. A total of 18 participants had thus inaccurately extended the work-life balance finding, accounting for 10.8% of all participants who provided responses and 90.0% of those who provided partially accurate responses. Additionally, an additional 2 responses included "increased productivity" in their description of the WISH evaluation findings.

Finally, 26 responses (15.6%) were coded as not applicable as they included no

description of findings or were judgement statements only.

# Table 18.

Frequency and samples of qualitative codes for accuracy of interpretation.

Code	Example Responses	Frequency
		( <i>n</i> = 167)
Inaccurate Statistics	<i>"It appears that by participating in the WISH program</i>	1.8%
(nominal numeric	women had about a 50% increase in happiness with work-	
value =-2)	life balance as well as income, while the comparison	
	group didn't see those kinds of improvements."	
Inaccurate Trend		0%
(nominal numeric		
value = $-1$ )		
Partially Accurate	"Training programs empower women. Results prove	12.0%
(nominal numeric	happiness vs drudgery, and 8% pay increase vs zero."	
value = $0$ )		
Subcode:	<i>"The findings were that women reported a higher quality"</i>	10.8%
Extend work-life	of life and a significant income increase."	
balance		
(Numeric Subcode	"Overall, women who participate in this program are	
=1)	happier and have a better life."	
Subcode:	"The study findings were that women who got to work	1.2%
Increased	from home vs women who did not get to participate in the	
productivity	program saw an increase in not only their satisfaction	
(Numeric subcode	with their work-life balance, but also an increase in their	
= 2)	incomes because they were happier and this could cause	
	production to increase. A happier employee, equals better	
	production."	
Accurate Trend	<b>Strong evidence strength condition:</b> <i>"Among program"</i>	46.7%
(Nominal numeric	participants, there was an increase in women reporting	
code = 1)	being extremely satisfied with work-life balance, but	
	among non-participants there was no increase. Among	
	participants, there was an overall increase in monthly	
	income compared to non-participants."	
	Weak evidence strength condition: "The program	
	seemed to nominally affect the women's satisfaction but	
	did increase the earnings though again by a nominal	
	amount."	

Code	Example Responses	Frequency
		( <i>n</i> = 167)
Accurate Statistics (Nominal numeric code = 2)	Strong evidence strength condition: "That women who participated in the WISH program gained in two ways. First the women in the program had a 10 to 55% increase in "extreme happiness in their work-life balance. The second benefit observed was an increase of 58% income. This is compared to women who did not avail themselves of the program." Weak evidence strength condition: "The WISH program	24.0%
	resulted in a 5% increase in the reporting metric of "extremely satisfied" by those who participated in the program, versus a lesser 2% increase in the control group. Additionally, those who participated in the WISH program resulted in an increase of 8% in the median monthly income of its participants, whereas there was no change in this metric for the control group."	
Not Applicable (Nominal numeric code = 3)	"Women's economic empowerment is central to realizing women's rights and gender equality. Women's economic empowerment includes women's ability to participate equally in existing markets; their access to and control over productive resources, access to decent work, control over their own time, lives and bodies; and increased voice, agency and meaningful participation in economic decision-making at all levels from the household to international institutions."	15.6%
	participating in WISH versus those who were not."	
Removed	"WOMEN PARTICIPATE THE LIFE BALANCE OF	n = 35
Kesponses	"agree"	

Accuracy codes were then assigned a nominal numerical value and merged with the quantitative data to explore the relationship between qualitative accuracy of interpretation and quantitative measures of motivation to elaborate level, data visualization approach, and evidence strength. Additionally, qualitative and quantitative measures of interpretation accuracy were compared to determine the extent to which these two measures correlated.

Chi-square test of independence analyses revealed no significant difference in qualitative interpretation accuracy based on low or high level of motivation to elaborate ( $\chi^2(3) = 6.514, p = 0.089$ ). Similarly, chi-square test of independence analyses showed no significant differences based on data visualization approach ( $\chi^2(3) = 4.582, p = 0.205$ ). However, significant differences in accuracy of interpretation were noted based on evidence strength ( $\chi^2(3) = 8.850, p = 0.031$ ). The association was moderately strong (Cohen, 1988; Cramer's V = 0.251). Using Agresti (2007) small table cutoff of greater than 2 in absolute value, standardized adjusted residuals indicate that weak evidence produced significantly higher than expected accurate statistical descriptions. Table 19 presents counts and standardized adjusted residuals for this finding.

## Table 19.

	Type of Evidence			
Interpretation Accuracy Code	Weak Evidence	Strong Evidence		
Inaccuracy Statistic	0	3		
	(-1.6)	(1.6)		
Partially Accurate	10	10		
	(0.4)	(-0.4)		
Accurate Trend	30	48		
	(-2.0)	(2.0)		
Accurate Statistic	25	15		
	(2.5)	(-2.5)		

Crosstabulation of qualitative interpretation accuracy and evidence strength.

Note. Adjusted residuals appear in parentheses below observed frequencies.

Responses were additionally coded for which of the four findings (WISH participants' work-life balance, control group's work-life balance, WISH participants' income, and control

group's income) were included in participants' descriptions. A total of 39 (23.4%) responses included descriptions of all four findings. Responses tended to include findings related to the WISH program (n = 144, 86.2%) while fewer responses included findings for the comparison group (n = 54, 32.3%). Those respondents who were presented weak evidence were slightly more likely to include descriptions of the control group than those who were presented strong evidence (29% vs. 25%); However, this trend was not statistically significant ( $\chi^2(1) =$ 1.403, p = 0.236). Respondents' descriptions of control group findings also included specific statistics less frequently than those who described WISH participants' findings. For work-life balance findings, respondents included specific statistics in 22.6% of descriptions of WISH participant findings and 8% of descriptions of control group findings. Similarly, in descriptions of income findings, respondents included specific statistics in 33.0% of descriptions of WISH participant findings and 11.6% of control group findings.

Another emergent theme in qualitative responses was inclusion of a judgement statement by 44 (26.3%) respondents, where participants expressed an appraisal of the evaluand. These were coded as negative, positive, or other. Only one participant included a negative appraisal of the evaluand, stating "It found that most women are not satisfied with their work/life balance". Positive appraisals of the evaluand included expressions that the program is effective, attributing outcomes to the program, or using positive adjectives to describe the outcomes (i.e., "better"). Examples of positive judgements of the evaluand described by participants are included in Table 20 below. One additional participant indicated that they were unable to make a judgement of the evaluand as they wanted to account for the cost of the program in their appraisal. No statistically significant differences were found in whether or not participants included a judgement statement based on evidence strength type ( $\chi^2(1) = 0.901$ , p = 0.342), data visualization approach

 $(\chi^2(1) = 0.067, p = 0.795)$ , or level of motivation to elaborate  $(\chi^2(1) = 1.488, p = 0.229)$ .

# Table 20.

*Example responses coded as positive judgements* (n = 42)*.* 

Evidence		
Strength		
Condition	Example Response	Frequency
Weak	• <i>"The women in the wish program faired better than the</i>	16
	comparison group in pay and satisfaction"	
	• "Wish Program is one of the best program . because most of	
	the women mentally and physically very depressed. this	
	program is such a wonderful opportunity, and their work	
	life balance to increased next level"	
	• "It was beneficial"	
Strong	• "The study findings showed that the WISH program really	26
C	works in helping with work-life balance and is beneficial and	
	helpful for women. The study showed that the women in the	
	program reported an increase of 45% in their satisfaction	
	with their work-life balance. This is a drastic improvement	
	compared to the comparison group, which did not see any	
	similar increase. Also, the study revealed that women who	
	participated in the program saw a 58% increase in their	
	median salary. This all shows that the program works and is	
	very beneficial to women's career."	
	• <i>"The results showed that work life balance improved by 30%"</i>	
	for participants in the program and participants income	
	increased by nearly 60%. Meaning the program is indeed	
	effective."	
	• "Women who had a better work life balance were happier	
	and made more money and WISH helped them achieve this."	

Finally, a total of 34 (20.4%) responses provided context for the findings by providing

descriptions of the WISH program's purpose (n = 14) or descriptions of the study design (n = 14)

20).

The one-week follow-up survey also collected participants' qualitative descriptions of the

WISH study findings. A total of 120 participants responded to the follow-up survey, however, 26

(21.7%) qualitative responses were removed due to off-topic or uninterpretable responses (see

Table 21 for sample responses). Of the remaining 94 responses, eight (8.5%) either stated they did not remember the findings or described a different study while 86 (91.5%) provide some description of the WISH study or its findings. These 86 responses were coded for accuracy and content using the same coding scheme as in the first section of this study (See Table 18).

## Table 21.

Code	Sample	Frequency
No memory	<ul> <li>"I don't remember them, please understand that I take a lot of studies, it's unreasonable to expect me to remember studies past 24 hours after doing them."</li> <li>"I don't remember this at all."</li> <li>"finding out covid and how women are handling it"</li> </ul>	8
Stability	<ul> <li>"I remember the study stating that women in the Wish program statistically were more likely to find stable employment, with a good work-life balance"</li> <li>"Women achieve greater financial stability by being able to work from home."</li> </ul>	3
Removed	<ul> <li>"The results section of the research paper is where you report the findings of your study based upon the information gathered as a result of the methodology you applied."</li> <li>"remarable one in our life"</li> <li>"human behavioral"</li> <li>"good and interesting"</li> </ul>	27

Frequency and samples of unique follow-up qualitative codes (N = 120).

Of the respondents who did remember the WISH evaluation (N=86), most provided responses that were partially correct (n = 29, 33.7%), accurate descriptions of general trends (n = 26, 30.2%), or provided contextual information only (n = 25, 29.0%). Only six individuals provided specific statistics in their summary of the findings, of which two responses included inaccurate statistics and four provided accurate statistics. Of those who provided partially correct information, 25 (86.2%) extended the work-life balance finding to alternative outcomes such as increased life satisfaction, job satisfaction, success, or happiness. For those that provided only contextual information, 22 (88.0%) mention the program's purpose and 3 (12.0%) noted the study design. Finally, six of the 26 individuals who provided accurate descriptions of general trends did not mention any specific outcome, but rather that the study findings were generally positive. For example, one participant wrote, "*The women who had WISH in their life performed better somehow*" while another described "*I think that women who went through the wish program were found to benefit over women who did not*". Accuracy of follow-up qualitative interpretations of WISH evaluation findings did not differ significantly based on evidence strength type ( $\chi^2(4) = 4.950, p = 0.292$ ), data visualization approach ( $\chi^2(4) = 3.092, p = 0.543$ ), or level of motivation to elaborate ( $\chi^2(4) = 1.049, p = 0.902$ ). Table 22 outlines how code frequencies in qualitative responses changed between the initial data collection and the one-week follow-up.

A total of nine responses (10.4%) mentioned all four WISH evaluation findings (WISH participants' work-life balance increase, control group's work-life balance increase, WISH participants' income increase, and control group's income). As with the original descriptions, one-week follow-up responses infrequently mentioned the control group's findings (n = 10, 11.6%). Of the 54 responses which noted WISH participants' findings, most (n = 25, 46.4%) included both work-life balance and income increases while 21 (38.9%) noted only the work-life balance finding.

Finally, seven (8.1%) of the follow-up responses included a judgement statement, all of which used positive adjectives to describe the evaluand and the study findings.

# Table 22.

	Initial Timepoint	<b>One-week Follow-up</b>
Code	( <i>N</i> = <b>167</b> )	( <i>N</i> = <b>86</b> )
Inaccurate Statistics	3	2
	(1.8%)	(2.3%)
Inaccurate Trend	0	0
	(0.0%)	(0.0%)
Partially accurate	20	29
	(12.0%)	(33.7%)
Accurate Trend	78	26
	(46.7%)	(30.2%)
Accurate Statistics	40	4
	(24.0%)	(4.7%)
Context Only	16	25
	(10.0%)	(29.0%)

Comparison of frequency of accuracy of initial and follow-up qualitative responses.

Analysis of Influence Pathway. As demonstrated in Figure 4, the hypothesized pathway from data visualization approach to elaboration traveled through data visualization engagement to involvement and motivation to elaborate. However, as summarized in the previous section, minimalist versus embellished data visualization approaches failed to produce differences in data visualization engagement, the first step in the hypothesized influence pathway. The question emerged, however, whether data visualization engagement could still produce differences in elaboration. As a result, the following exploratory analyses were conducted to explore this relationship.

First, the relationship between data visualization engagement and elaboration as seen through global attitude change was explored through hierarchical multiple regression to test the interaction between data visualization engagement and evidence strength. Table 23 below summarizes the descriptive statistics for the two continuous variables.

#### Table 23.

Descriptive statistics for continuous variables included in regression analyses.

Variable	N	Mean	SD	Min	Max	Skew	Kurtosis
Global Attitude Change	201	0.14	0.50	-2.00	2.40	0.36	4.45
Engagement	201	4.90	1.14	1.47	7.00	-0.54	-0.10

There was independence of residuals (Durbin-Watson = 1.944). Engagement scores were centered to reduce multicollinearity. Examination of studentized residual plots and partial regression plots confirmed linear relationships between global attitude change and centered engagement scores. There was homoscedasticity, as assessed by visual inspection of a plot of studentized residuals versus unstandardized predicted values. Pearson's correlations indicated no multicollinearity (Pearson's |r|<0.7) and the minimum Tolerance statistic was 0.610.

A total of 4 outliers were identified as having standardized residuals greater than 3 standard deviations. However, as no case had a leverage point greater than 0.2, these were retained in the analyses. Similarly, no cases had a Cook's Distance value above 1 (maximum Cook's Distance = 0.23). Examination of P-P Plot indicated approximately normally distributed residual errors.

The addition of the interaction term between engagement scores and evidence strength led to a statistically significant increase in  $R^2$  of 0.020, F(1, 197) = 4.009, p = 0.047. The full model of engagement score, evidence strength, and the interaction term significantly predicted global attitude change,  $R^2 = 0.041$ , F(3, 197) = 2.804, p = 0.041, adjusted  $R^2 = 0.026$ . Examination of coefficients indicated a significant interaction between engagement

scores and evidence strength t(197) = -2.002, p = 0.047, B = -0.125, 95% CI (-0.248 to -0.002),

 $\beta = -0.179$ . Table 24 below summarizes the regression results.

# Table 24.

Moderation effects of evidence strength on engagement scores predicting global attitude change.

Step	Variable	r	$R^2$ Change	В	SE <sub>B</sub>	β
1	Engagement Score	-0.040	0.002	0.030	0.039	0.070
2	<b>Evidence Strength</b>	0.140*	0.020*	0.140*	0.069	0.141*
	(Weak = 0;					
	Strong=1)					
3	Interaction Term	0.097	0.020*	-0.125	0.062	-0.179
	(Engagement Score					
	x Evidence					
	Strength)					
(Constant)				0.065	0.049	

\*p < 0.05; \*\*p < -.01; p < 0.001; Cumulative  $R^2 = 0.041$ ; Adjusted  $R^2 = 0.026$ ; Engagement is 'centered' to a mean of 4.90.

Figure 14 illustrates the interaction between engagement and evidence strength on

predicted global attitude change utilizing the following regression equation:

# Equation 1.

Regression predicting global attitude change.

 $\hat{Y} = 0.065 + 0.03$ (Engagement Score) + 0.14(Evidence Strength Level)

- 0.125 (Engagement Score x Evidence Strength Level)

When presented strong evidence, individuals with low engagement with the data

visualization demonstrate greater global attitude change. Conversely, when presented weak

evidence, individuals with greater engagement with the visual demonstrate greater global attitude

change.

Figure 14.

Predicted global attitude change based on centered engagement scores and evidence strength.



*Note*. Figure utilizes engagement scores  $\pm 1$  standard deviation from the mean for illustrative purposes. Figure is not a complete description of the data.

Similarly, the relationship between data visualization engagement and elaboration as seen through evaluand-specific attitude change was explored through hierarchical multiple regression to test the interaction between data visualization engagement and evidence strength. Table 25 below summarizes the descriptive statistics for the additional continuous variables.

#### Table 25.

Descriptive statistics for evaluand-specific attitude change.

Dependent Variable	N	Mean	SD	Min	Max	Skew	Kurtosis
Specific Attitude Change	201	0.20	0.57	-1.20	3.20	1.27	5.35

There was independence of residuals (Durbin-Watson = 2.049). Engagement scores were centered to reduce multicollinearity. Examination of studentized residual plots and partial regression plots confirmed linear relationships between evaluand-specific attitude change and centered engagement scores. There was homoscedasticity, as assessed by visual inspection of a plot of studentized residuals versus unstandardized predicted values. Pearson's correlations

indicated no multicollinearity (Pearson's |r|<0.7) and the minimum Tolerance statistic was 0.610.

A total of 3 outliers were identified as having standardized residuals greater than 3 standard deviations. However, as no case had a leverage point greater than 0.2, these were retained in the analyses. Similarly, no cases had a Cook's Distance value above 1 (maximum Cook's Distance = 0.13). Examination of P-P Plot indicated approximately normally distributed residual errors.

The addition of the interaction term between engagement scores and evidence strength led to a statistically significant increase in  $R^2$  of 0.024, F(1, 197) = 5.176, p = 0.024. The full model of engagement score, evidence strength, and the interaction term significantly predicted specific attitude change,  $R^2 = 0.088$ , F(3, 197) = 1.879, p < 0.001, adjusted  $R^2 = 0.074$ .

Examination of coefficients indicated a significant interaction between engagement scores and evidence strength t(197) = -2.275, p = 0.024, B = -0.157, 95%CI (-0.294 to -0.021),  $\beta = -0.198$ . Table 26 below summarizes the regression results.

#### Table 26.

Moderation effects of evidence strength on engagement scores predicting specific attitude change.

Step	Variable	r	$R^2$ Change	В	SE <sub>B</sub>	β
1	Engagement Score	-0.141*	0.020*	-0.010	0.043	-0.020
2	<b>Evidence Strength</b>	0.208	0.044**	0.238**	0.077	0.210**
	(Weak = 0;					
	Strong=1)					
3	Interaction Term		0.024*	-0.157*	0.069	-0.198*
	(Engagement Score					
	x Evidence					
	Strength)					
(Constant)			0.079	0.055		
* $P_{1}^{(2)} = 0.05$ ** $P_{2}^{(2)} = 0.001$ Cymulative $P_{2}^{(2)} = 0.088$ A diveted $P_{2}^{(2)} = 0.074$ Engagement is						

\*p < 0.05; \*\*p < -.01; p < 0.001; Cumulative  $R^2 = 0.088$ ; Adjusted  $R^2 = 0.074$ ; Engagement is 'centered' to a mean of 4.90.

Findings result in the following regression equation to predict evaluand-specific attitude change:

# **Equation 2.**

Regression equation predicting evaluand-specific attitude change.

 $\widehat{Y} = 0.079 - 0.01$ (Engagement Score) + 0.238(Evidence Strength Level)

- 0.157(Engagement Score x Evidence Strength Level)

As demonstrated in Figure 15 below, individuals show similar changes in specific attitude scores regardless of the degree to which they found the data visualization engaging when evidence is weak. However, when evidence is strong, engagement plays a significant role in predicting specific attitude change. Those that found the data visualization highly engaging demonstrated greater changes in evaluand-specific attitudes than those that reported low engagement scores.

### Figure 15.

Predicted evaluand-specific attitude change based on centered engagement scores and evidence strength.



*Note*. Figure utilizes engagement scores  $\pm 1$  standard deviation from the mean for illustrative purposes. Figure is not a complete description of the data.

#### DISCUSSION

Phase 1 of the research explored how experience of data visualization and key interpretation and attitudinal outcomes differed based on motivation to elaborate, data visualization approach, and evidence strength; as such, exploring the initial steps of the hypothesized evaluation influence pathway. Overall, the hypothesized evaluation influence pathway from evaluation product (data visualization) to attitude change was partially supported. Additionally, findings indicate that pre-existing motivation to elaborate plays a significant role in both experience and outcomes, some of which are further dependent on the strength of the evaluation evidence. Findings also indicate that the data visualization approach did not make a discernable difference in either experiences of data visualizations or their outcomes.

One question explored through this research was whether evaluation product design could be informed by the Elaboration Likelihood Model (ELM) to support the formation of evaluation influence pathways. The initial step of the hypothesized evaluation influence pathway – that from data visualization design to engagement – was not supported (hypotheses 1.2 and 1.3). These findings have implications for data visualization research, as proponents of embellished visualization approaches argue that such conventions are more engaging (e.g., Bateman et al., 2010; Byrne, Angus, & Wiles, 2016). The current findings suggest that engagement does not have to do with how the visualization is presented but rather characteristics of its audience, as discussed further below. Future research into data visualization approaches should consider motivation to elaborate as a factor in experience of data visualization.

However, despite this initial misstep in the hypothesized model, exploratory findings confirmed the hypothesized steps from engagement to elaboration and attitude change in the evaluation influence pathway. Those that were highly engaged in the data visualization

differentiated the degree to which their attitudes towards the evaluand changed based on the strength of the evaluation findings – providing evidence for underlying elaboration processes. Interestingly, similar trends were seen for those who did not engage with the data visualization for attitude change towards women empowerment programs more globally. It is unclear why these two types of attitudes showed reverse trends and future research should further explore the underlying mechanisms of these relationships.

Previous research had not explored the connection between motivation to elaborate and experiences of data visualization. The findings from Phase 1 of this research, however, provide evidence that pre-existing motivation to elaborate plays a significant role in the experience of data visualization. In particular, the extent to which participants found data visualization engaging and were satisfied with the visualization differed significantly based on level of motivation to elaborate, such that those who were highly motivated to elaborate reported more engagement and more satisfaction with the visualizations. It is not surprising that those who are not interested in learning more about a program would not find an evaluation report engaging or satisfying. A limitation of this study is that all participants were asked to review the evaluation report, limiting external validity of the findings. It is possible that outside an experimental setting, those who do not find such tasks enjoyable due to lack of motivation to elaborate may not read an evaluation report at all. As such, evaluators may want to explore options for ensuring that those low in motivation to elaborate read evaluation reports or are otherwise exposed to the evaluation findings, especially given that these individuals are also more likely to show significant attitude change, as discussed further below.

Pre-existing motivation to elaborate also played a role in how long participants examined the visualization, though this was further dependent on evidence strength, but not data

visualization approach. Thus, hypothesis 1.1 was only partially supported. When evidence was weak, those who were highly motivated to elaborate took longer to examine the information than when the evidence was strong. Trends indicated the reverse was true for those low in motivation to elaborate; however, the difference in duration spent with the visual between weak and strong evidence was not statistically significant for those low in motivation to elaborate. There is insufficient evidence to explain why those who were highly motivated to elaborate took longer to examine weak evidence. It is possible that these individuals were surprised by the findings given that they also tended to have more positive initial attitudes towards the evaluand, and this cognitive dissonance prompted a closer examination of the findings. Phase 3 of this research explores how prior expectations of results impact visualization experience and outcomes. In general, further research is needed to determine the underlying cause for the differences in how long participants spend reading through evaluation reporting.

Regardless, there was no evidence that minimalist data visualizations took less time to examine compared to embellished graphs, as claimed by proponents of minimalist approaches (e.g., Cleveland & McGill, 1985; Gillan & Richman, 1994, Tufte, 1983). Phase 1 of this research did not replicate these findings. Instead, duration of examination had more to do with characteristics of the audience and the content of the visualization, which had not been examined in previous data visualization research. Future research comparing embellished and minimalist data visualization approaches should include these factors to determine if any differences beyond those examined here exist in data visualization experience or outcomes based on visualization convention. For evaluators, this means flexibility in the data visualizations they can use in their evaluation products, as neither design approach showed significant practical advantages.
Stakeholders tend to have multiple demands on their time, and it may be helpful to understand how audiences of evaluation reporting may differ in how long they take to review such a report. Phase 1 findings of this research suggest that evaluators may want to include considerations of their audience's pre-existing motivation to elaborate when anticipating the demand on the stakeholder's time a report will place. If evaluation findings are weak or surprising, evidence indicates that evaluators may need to anticipate stakeholders taking longer to examine the materials.

Motivation to elaborate also played a significant role in the interpretation and attitudinal outcomes included in Phase 1. The following paragraphs describe impact of motivation to elaborate on accuracy of interpretation, perceptions of the evaluation study's trustworthiness, attitude change, and memorability as well as implications for evaluation practice and data visualization research.

Accurate interpretation of evaluation findings is critical, as it could impact the ability of an evaluation to achieve social betterment. Inaccurately interpreted findings could lead to poorly informed decision-making or even misinformation. Similarly, proponents of minimalist approaches to data visualization emphasize that the purpose of data visualization is to accurately convey findings, thus making accuracy of interpretation a key outcome of quality data visualization (Tufte, 1983). There have been mixed findings regarding the effectiveness of embellished graphs to produce accurate data interpretation (Bateman et al., 2010; Fisher, 2010). Based on Phase 1 findings, there were no differences in accuracy of data interpretation based on data visualization approach. Instead, the level of motivation to elaborate differentiated those who responded to interpretation questions with high accuracy. Specifically, those who had low levels of motivation to elaborate provided more accurate interpretations of the data than those who

were highly motivated to elaborate. This has significant practical implications for evaluation, as it suggests that, at least among the public, those with a greater stake in the evaluand may be less effective at interpreting evaluation findings, regardless of their previous experience with data visualization or the data visualization approach utilized. This finding suggests that evaluators may want to incorporate practices that assess stakeholders' accurate understanding of evaluation findings.

It is unclear from the findings why those who were low in motivation to elaborate had more accurate interpretations. One possibility is that it depends on how accuracy is measured: qualitative responses did not replicate the main effect of motivation to elaborate; instead, evidence strength played a significant role in the accuracy of qualitative descriptions of the evaluation findings. Those who were presented weak findings were more likely to incorporate accurate statistics into their descriptions of the findings compared to those who were presented strong findings. Qualitative findings also indicate that most participants tended to focus on general trends while the quantitative measure of interpretation accuracy focused more heavily on specific statistics. Research has found that the persuasiveness of data visualization depended on alignment of finding with pre-existing beliefs, such that those with either neutral or aligned preexisting beliefs were more likely to be persuaded by figures over tables as figures emphasized general trends. Conversely, those with misaligned pre-existing beliefs were more convinced by tables over figures as these allow for closer examination of counterfactuals (Mason & Azzam, 2018; Pandey et al., 2014). Similar underlying mechanisms may be at work in the current research, as those who were low in motivation to elaborate also tended to have less positive preexisting attitudes towards the evaluand. Further research is needed to establish why those who are highly motivated to elaborate do not accurately interpret evaluation findings.

In addition, motivation to elaborate also played a significant role in attitude change specific to the evaluand. In an inverse from hypothesis 2.4, findings indicate that those who were high in motivation to elaborate tended to change their attitudes less than those who were low in motivation to elaborate. Moreover, findings that attitude change did not depend on evidence strength suggests that neither group engaged in significant elaboration. This finding is at odds with previous research on the Elaboration Likelihood Model (ELM), where those who were highly motivated to elaborate tended to show significant attitude change when presented with strong evidence (e.g., Haugtvedt et al., 1995; Petty & Cacioppo, 1986; Wang, Wang, & Farn, 2009). This finding carries significant implications for evaluation. On the one hand, social programs often need support not only from those who would benefit from the program, but others as well - including decision-makers and, for publicly funded programs, support from the public. For example, in some places, there are not enough parents with young children to pass legislation that funds critical early childhood programs – support from non-parents is needed in these cases. Findings from Phase 1 suggest that evaluation findings may be an effective tool for persuading non-parents – who would likely be low in motivation to elaborate – to support such programming. However, evaluation reporting tends to be geared towards primary stakeholders – defined as those who have a stake in the evaluation (Weiss, 1998). These individuals are likely to be high in motivation to elaborate as the evaluand is likely to be highly relevant to them. Phase 1 findings suggest that these individuals are less likely to change their attitudes based on evaluation findings. The implication is two-fold: Focusing evaluation reporting on audiences that are likely high in motivation to elaborate (key stakeholders) may (1) be ineffective and (2) limit the possible impact the evaluation could have on a broader audience. Indeed, this may play a role in why Fleischer and Christie (2009) found limited reports of evaluation use.

However, one possible reason for the attitude change findings is that motivation to elaborate was based on relevance of the evaluand. It may be that other factors which affect motivation to elaborate may produce different results. For example, one factor that affects motivation to elaborate is involvement, where the individual has some vested interest or responsibility, such as to make a decision, based on the persuasive argument (Crano & Sivacek, 1982; Petty & Cacioppo, 1986). This factor may be more apt to be found among evaluation stakeholders such as decision-makers and funders and may result in different processing of the evaluation findings. This possibility is further explored in Phase 3 of this research.

It is also important to consider that previous research into the ELM has focused on arguments that do not include scientific evidence (e.g., "Increased tuition will pay for improvements to university library" vs. "Independent research showed that increasing tuition tends to result in improved resources for students"). It may be that evaluation findings are processed differently from the types of arguments typically used in ELM research. One possible reason for this is that the evaluation may serve a different role in the persuasive process – such as a persuasive argument, a peripheral cue, or a biasing factor (Petty & Cacioppo, 1986). Any of these three roles would be consistent with attitude change that is not dependent on the evidence strength. An alternative explanation may be that there is a ceiling effect to attitudes towards the WISH program, and those that were highly motivated to elaborate started out with more positive attitudes towards the evaluand could not have even more positive attitudes; however, this is unlikely as average attitudes after the intervention were below the top score of the scale.

Interestingly, even though global attitudes towards programs similar to the evaluand significantly increased, none of the explored factors (motivation to elaborate, data visualization approach, and evidence strength) accounted for this change. As the change was smaller than that

of evaluand-specific attitudes, it may be that the Phase 1 study did not have sufficient power to detect effects. Further research is needed to understand the factors that impact global attitude change.

Participant's trust of the evaluation findings was dependent on both motivation to elaborate and the strength of the evidence presented. Those low in motivation to elaborate tended to not trust weak findings and to trust strong findings, while those high in motivation to elaborate did not differentiate perceptions of trustworthiness based on evidence strength. This finding is the reverse of hypothesis 2.2, which anticipated that those high in motivation to elaborate would differentiate trust based on evidence strength, consistent with the Elaboration Likelihood Model (ELM). However, previous research on the ELM has focused on attitude change specific to the subject of the persuasive message and did not explore whether participants trusted the arguments provided in the message (e.g., Areni, 2003; Petty & Cacioppo, 1986). As a result, it is possible that perceptions of trustworthiness are formed differently than attitudes and do not follow the same processing as the ELM. The finding does indicate that evaluators should be especially cautious when presenting weak findings and, to the extent possible, address counterfactuals to strengthen their conclusions, or risk some audiences mistrusting the evaluation study.

Finally, the first phase of this research explored memorability, assessed as both recall and accuracy of memory. All three factors – motivation to elaborate, data visualization approach, and evidence strength – played a role in recall, with data visualization approach having the greatest impact. Bateman and colleagues (2010) had argued that three possible reasons why embellished graphs may result in better recall: (1) Presence of an image provides additional encoding in memory which can improve recall, (2) the embellished images are very different from one another, while the plan bar charts all have similar appearances, and (3) user's emotional response

to the imagery in embellished charts could help anchor the chart details in a viewer's memory. In contrast to such suggestions by advocates of embellished graphs, minimalist graphs outperformed embellished graphs in accurate recall in the current research. Those presented minimalist graphs were able to accurately identify the visualizations they had been presented – both in terms of visualization approach and evidence strength – while those who were presented embellished graphs were able to recall the correct evidence strength, but not how the visualization looked.

The positive news for evaluators is that both visualization conventions resulted in most participants recognizing the evaluation findings. However, given that Mark and Henry's (2004) Evaluation Influence model poses that there are influence pathways, and these pathways may take significant time, it may be beneficial for evaluators to use minimalist approaches to visualizing data to support accurate recall.

The second most impactful factor on recall was evidence strength, such that those presented weak evidence were more likely to have no recollection while those presented strong evidence recognized both the correct findings and visualization. This finding further emphasizes the need for strong evaluation findings as a key factor which may have a role in evaluation influence pathways.

As predicted in hypothesis 2.3, motivation to elaborate played a role in recall; however, findings were opposite than those predicted – those who were low in motivation to elaborate had better recall than those who were highly motivated to elaborate. This finding further emphasizes the possible role for non-stakeholders in evaluation influence pathways. As this group has been found in the current research to have greater attitude change and better recall, they may be poised to take further steps in the evaluation influence pathway – perhaps by becoming a champion for

the program or engaging in individual behaviors that support the program (Mark & Henry, 2004). This possibility is further explored in Phase 2 of this research.

# CHAPTER 3 PHASE II: EXPERIMENTAL EXPLORATION OF EFFECTS OF ELABORATION AND DATA VISUALIZATION ON ATTITUDE STRENGTH AND BEHAVIOR

The second phase of this research builds upon the first by exploring how data visualization approaches and elaboration on evaluation evidence affect attitude strength and donation behaviors. Attitudes are one of three key factors that influence behavioral intent (See Figure 3; Ajzen, 1991). Additionally, previous research has linked greater elaboration with greater attitude strength, defined as attitude resilience, attitude persistence, and behavioral intent (e.g., Crano, 1995; Petty & Cacioppo, 1986). If different levels of data visualization engagement result in stronger attitudes, this may lead to greater behavioral intent and a higher likelihood of actual behaviors. Building on the results of Phase 1, Phase 2 explores whether more engaging visualizations are associated with higher likelihood of elaboration and whether these result in greater attitude strength and donation behaviors. Understanding this relationship can bring greater understanding to the creation of evaluation influence pathways. Presented below are hypotheses for the roles of motivation to elaborate, data visualization approach, and evidence strength in participants' attitude strength and donation behaviors; Please note that the following presents both original hypotheses and hypotheses that were updated based on the findings of Phase 1 of this research.

- 3. Does elaborating on evaluation findings result in greater likelihood of evaluation influence pathways?
  - a. *Original Hypothesis 3.1:* Those who are high in motivation to elaborate will demonstrate higher attitude persistence and resilience than those who are low in motivation to

elaborate. This trend will be more pronounced for those viewing embellished visualization compared to those viewing minimalist visualizations.

*Updated Hypothesis 3.1:* Those low in motivation to elaborate will demonstrate higher attitude persistence and resilience than those who are low in motivation to elaborate. Additionally, those who highly engage with their data visualization will demonstrate evaluand-specific attitude persistence and resilience when evidence is strong but not when evidence is weak.

b. Original Hypothesis 3.2: Those who are high in motivation to elaborate will demonstrate greater intent to support the evaluand than those who are low in motivation to elaborate. This trend will be more pronounced for those viewing embellished visualization compared to those viewing minimalist visualizations.

*Updated Hypothesis 3.2:* Those low in motivation to elaborate will demonstrate greater intent to support the evaluand than those who are high in motivation to elaborate. Additionally, those who highly engage with their data visualization will demonstrate greater intent to support the evaluand when evidence is strong but not when evidence is weak.

c. *Original Hypothesis 3.3:* Those who are high in motivation to elaborate will donate more funds to support the evaluand than those who are low in motivation to elaborate. This trend will be more pronounced for those viewing embellished visualization compared to those viewing minimalist visualizations.

*Updated Hypothesis 3.2*: Those low in motivation to elaborate will donate more funds to support the WISH program than those who are high in motivation to elaborate. Additionally, those who highly engage with their data visualization will donate more

funds to support the WISH program when evidence is strong but not when evidence is weak.

Together, these hypotheses explore the full hypothesized pathway to evaluation influence as noted in Figure 4.

**Design.** As in the previous study, Phase 2 used a 2 (motivation to elaborate: low vs. high) x 2 (data visualization approach: embellished vs. minimalist) x 2 (evidence strength: weak vs. strong) between-subjects experimental design.

#### METHODS

**Recruitment.** As in the first phase of this research, the Phase 2 study recruited participants from Amazon Mechanical Turk (MTurk), a crowdsourcing platform. The study participants were limited to Amazon MTurk workers who reside in the United States to minimize potential confounding factors and the risk of language barriers as the study was conducted in American English.

**Procedure.** Like the previous study, Phase 2 used a simulation of a program evaluation report, with eight stages to the study procedure (See Figure 16). First, participants had their motivation to elaborate assessed. Second, participants were presented with a short description of the WISH program and asked a short survey about their attitudes towards women empowerment programs generally and the described WISH program specifically. Third, participants were randomly presented with one of four presentations of the evaluation findings. Fourth, participants were again asked about their general and specific attitudes toward the WISH program and women empowerment programs. Following the attitudes survey, participants were asked a series of questions regarding their willingness to engage in behaviors that would be supportive of the WISH program. Upon completion of the behavioral intent survey, participants were given the

opportunity to donate their study earnings to support the WISH program's funding. Sixth, participants reported on their engagement with the visual. Finally, participants completed the demographic questions. From this point, the sample was split into two groups through random assignment.

The first group was assigned to the attitude resilience survey, which presented them with a message in opposition of the WISH program (it is more expensive to implement than other similar programs), and participants' attitudes towards the WISH program were measured a final time.

For the second group, a week delay was introduced before gathering data regarding attitude persistence. A week delay was selected in order to maximize retention while ensuring some time has passed that could dilute attitudes. Participants were re-contacted through Amazon MTurk, and a description of the WISH program was provided again. Finally, the subgroup assigned to the attitude persistence measure filled out the post-test attitudes measure. All participants were debriefed at the conclusion of the procedure.

Participants were paid up to \$4.00 in incentives, with \$2.40 upon completion of the first survey, and an additional \$1.60 upon completion of either the resilience or persistence questionnaires.

#### Figure 16.

Eight stages of Phase 3 Procedure



*Motivation to Elaborate*. The approach to measuring motivation to elaborate replicated the approach used in Phase 1.

*Scenario.* The same scenario of a hypothetical evaluation of the WISH program was utilized in Phase 2 as in Phase 1. However, for the purposes of Phase 2, it was not revealed that the scenario was of a hypothetical program to ensure accurate capture of donation behavior. The simulation approach was revealed to participants during debrief at the completion of the study.

Attitude Measure. Phase 2 reused the attitude measure from Phase 1.

Data Display. Phase 2 reused the data display materials from Phase 1.

*Evidence Strength.* Phase 2 reused the same evaluation finding effect sizes as in Phase 1. *Engagement Measure.* Phase 2 used the same engagement measure as Phase 1.

**Behavioral Intent Survey**. In order to assess behavioral intent, participants were asked four survey items regarding the extent to which they would engage in a series of behaviors in support of the WISH program (e.g., tell a friend about the Program, voting on a ballot measure to provide funding to the Program) based on a 7-point Likert scale (1 = Not at all, 7 = Definitely). The Behavioral Intention Survey items and reliability analyses are provided in Appendix J. Behavioral intent scores were calculated based on averaging the four items.

*Measure of Donation Behavior*. Replicating an approach used by Blueshtein (2018) to measure pro-social behaviors, participants were also given the option to donate a portion of their earnings from the task to support the WISH program through two questions. First, participants were asked a dichotomous question whether they wish to donate any portion of their HIT earnings. For those who answer "yes," participants were asked how much of their earnings they wished to donate as a continuous numeric variable with options ranging from \$0.00 to \$2.40 (the total task payment for initial survey).

*Demographics.* The same demographic questions were used as in the previous study.

Attitude Resilience. To measure attitude resilience, participants were presented with the following information in opposition to the WISH program that is unrelated to the evaluation findings: "A recent article by a local newspaper reports that the WISH program is significantly more expensive than other programs that address economic opportunity and work-life balance for women". After presentation of this information, the attitudes measure was administered a final time. A change score was calculated based on comparison with attitudes scores on the immediate post-intervention attitude measure. Those with no to small change have high attitude resilience while those with a large change have low attitude resilience.

Attitude Persistence. Finally, to assess attitude persistence, the attitude measures were readministered after a week period of delay. As in the previous phase of research, participants were assigned a worker qualification in the Amazon MTurk portal indicating their previous participation in the study and an email was sent using Amazon MTurk's API and Python programming (see Appendix B for full email text). The study oversampled to account for attrition rates, and attrition across conditions was analyzed to ensure equivalent groups. Persistence was calculated based on the difference score between the immediate postintervention attitude measure and the attitude measure after a week delay. Those with no to small change have high attitude persistence while those with a large change have low attitude persistence.

#### RESULTS

**Population.** The study recruited 361 participants from Amazon Mechanical Turk (MTurk), a crowdsourcing platform. A total of 21 participants failed to match on gender demographics from their reported gender on Amazon MTurk and were removed from the analyses for a final total

sample of 340. To reduce research fatigue and minimize risk of biased results due to order effects, the sample was split into two: 150 participants completed an attitude resilience measure immediately upon completion of the main portion of the study while a separate 211 participants were invited to participate in a follow-up survey after a week's delay to measure attitude persistence. Of these, 132 responded for a 62.6% retention rate. There were no significant differences in retention between conditions ( $\chi^2(3) = 1.075$ , p = 0.783). This section presents the demographics for both the overall sample and each subsample.

As can be seen in Table 27, while the overall sample contained roughly equal representation from male and female participants, males were slightly more represented in the resilience subsample. Additionally, males were significantly more likely to be retained in the persistence subsample ( $\chi^2(2) = 6.070$ , p = 0.048). Study condition did not differ significantly based on gender for the overall sample ( $\chi^2(9) = 9.720$ , p = 0.374), the resilience subsample ( $\chi^2(6) = 6.842$ , p = 0.336), the recruited persistence subsample ( $\chi^2(6) = 5.083$ , p = 0.533), nor the retained persistence subsample ( $\chi^2(3) = 1.490$ , p = 0.685).

#### Table 27.

Gender demographics for	Phase 2 research	sample and	subsamples.

Sample	N	Male	Female
Overall Sample	340	166 (48.8%)	172 (50.6%)
Resilience Subsample	150	83 (55.3%)	66 (44.0%)
Persistence Subsample Recruited	211	84 (39.8%)	126 (59.7%)
Persistence Subsample Retained	132	60 (45.5%)	72 (54.5%)
		71.4% retained	57.1% retained

Most respondents in the initial sample were either 25-34 years old (41.5%) or 35-44 years old (23.5%, See Table 28). No participant reported being over 74 years old. Based on Pearson

Chi-Square analyses, there were no significant differences in age group based on study condition in the overall sample ( $\chi^2(15) = 12.155, p = 0.667$ ), resilience subsample ( $\chi^2(15) = 15.963, p = 0.384$ ), recruited persistence subsample ( $\chi^2(15) = 11.442, p = 0.698$ ), nor retained persistence subsample ( $\chi^2(15) = 15.164, p = 0.440$ ).

#### Table 28.

Phase 2 participants by age groups.

Age (years)	18 to 24	25 to 34	35 to 44	45 to 54	55 to 64	65 t0 74
Overall Sample ( $N = 340$ )	5.0%	41.5%	23.5%	15.9%	11.2%	2.9%
Resilience Subsample ( $N = 150$ )	4.0%	38.0%	26.7%	16.0%	12.0%	3.3%
Persistence Subsample Recruited ( $N = 211$ )	5.2%	43.1%	20.4%	18.0%	10.4%	2.8%
Persistence Subsample Retained ( $N = 132$ )	5.3%	40.9%	18.2%	19.7%	12.9%	3.0%

Additionally, most participants in the overall sample (40.6%) had completed a bachelor's degree and an additional 26.2% had completed a master's degree or Professional Degree (See Table 29). Similar trends held for the subsamples. Based on Pearson Chi-Squared analyses, there were no significant differences in education level across conditions in the overall sample ( $\chi^2(18) = 18.920, p = 0.397$ ), resilience subsample ( $\chi^2(18) = 22.680, p = 0.203$ ), recruited persistence subsample ( $\chi^2(15) = 21.228, p = 0.130$ ), nor retained persistence subsample ( $\chi^2(9) = 13.103, p = 0.158$ ).

#### Table 29.

Participants'	highest l	level of	competed	education.
---------------	-----------	----------	----------	------------

			Persistence	Persistence	
	Overall	Resilience	Subsample	Subsample	
	Sample	Subsample	Recruited	Retained	
N	340	150	211	132	
Some High School	0.6%	0.7%	0.5%	0.0%	
High School Diploma/	12 50/	10 70/	14 20/	12 10/	
GED	15.5%	10.7%	14.2%	12.170	
Associate's Degree/	17 60/	16.00/	19 00/	11 40/	
Some College	17.0%	10.0%	18.0%	11.4%	
Bachelor's Degree	40.6%	40.0%	42.7%	47.7%	
Master's Degree/	26.20/	20.00/	24.20/	28.80/	
Professional Degree	20.2%	50.0%	24.2%	20.8%	
Doctorate/PhD	1.2%	2.0%	0.5%	0.0%	

Demographic questions also explored participants' familiarity with data visualization through three questions. Overall, results indicate that participants tended to see visualized data (M = 3.95, SD = 1.18) and use data displayed in charts to make decisions (M = 3.54, SD = 1.22)on a weekly basis but did not typically create data visualizations themselves (M = 2.97, SD =1.45). In the persistence subsample, there were no significant differences in mean responses to these questions between those that completed the follow-up survey and those that did not (t(209)= -1.468, p = 0.143). One-way ANOVA analyses revealed no differences in data visualization experience between study conditions in the overall sample (F(3, 336) = 1.312, p = 0.271), the resilience subsample (F(3, 146) = 5.153, p = 0.169), the recruited persistence subsample (F(3, 207) = 0.391, p = 0.70), nor the retained persistence subsample (F(3, 128) = 0.512, p = 0.674). **Attitude Change Analyses.** The first set of analyses explores whether the relationships between data visualization engagement, evidence strength, and attitude change (both global and specific) are replicated from Phase 1. Hierarchical multiple regression was utilized to explore these effects. Table 30 summarizes the descriptive statistics for the continuous variables.

## Table 30.

Descriptive statistics of attitude change and engagement scores for overall sample.

Variable	N	Mean	SD	Min	Max	Skew	Kurtosis
Global Attitude Change	340	0.04	0.66	-2.80	5.20	1.14	14.69
Specific Attitude Change	340	0.06	0.65	-3.20	2.40	-0.98	5.50
Engagement	340	5.32	1.19	2.05	8.00	-0.12	-0.50

The initial hierarchical multiple regression analysis utilized global attitude change as the dependent variable. There was independence of residuals (Durbin-Watson = 2.171). Examination of studentized residual plots and partial regression plots confirmed linear relationships between global attitude change and centered engagement scores. There was homoscedasticity, as assessed by visual inspection of a plot of studentized residuals versus unstandardized predicted values.

A total of 8 outliers were identified as having standardized residuals greater than 3 standard deviations. However, as no case had a leverage point greater than 0.2, these were retained in the analyses. Similarly, no cases had a Cook's Distance value above 1 (maximum Cook's Distance = 0.18). Examination of P-P Plot indicated approximately normally distributed residual errors.

The addition of the interaction term between engagement scores and evidence strength did not lead to a statistically significant increase in  $R^2$ , F(1, 336) = 0.07, p = 0.113. The full model of engagement score, evidence strength, and the interaction term significantly predicted

global attitude change,  $R^2 = 0.039$ , F(3, 336) = 4.530, p = 0.004, adjusted  $R^2 = -0.030$  (See Table

31).

## Table 31.

Regression results predicting global attitude change.

Step	Variable	r	$R^2$ Change	В	SE <sub>B</sub>	β
1	Engagement Score	0.115*	0.013	0.104*	0.041	0.187*
2	Evidence Strength	0.142**	0.018	0.180*	0.071	0.136*
	(Weak $= 0;$					
	Strong=1)					
3	Interaction Term		0.007	-0.095	0.060	-0.117
	(Engagement Score					
	x Evidence					
	Strength)					
(Cons	stant)			-0.051	0.050	

\*p < 0.05; \*\*p < -.01; p < 0.001; Cumulative  $R^2 = 0.039$ ; Adjusted  $R^2 = -0.030$ ; Engagement is 'centered' to a mean of 5.32.

The resulting regression equation predicting global attitude change is as follows:

## **Equation 3.**

Regression equation predicting global attitude change.

 $\hat{Y} = -0.051 + 0.104$  (Engagement Score) + 0.180 (Evidence Strength Level)

- 0.095(Engagement Score x Evidence Strength Level)

As can be seen in Figure 17, higher levels of engagement predicted slightly greater global

attitude change. Conversely, as illustrated in Figure 18, strong evidence predicted smaller

attitude change than weak evidence.

# Figure 17.

Effects of engagement on predicted global attitude change.



Figure 18.

Effects of evidence strength on predicted global attitude change.



The second hierarchical multiple regression analysis utilized evaluand-specific attitude change as the dependent variable. There was independence of residuals (Durbin-Watson = 1.961). Examination of studentized residual plots and partial regression plots confirmed linear relationships between evaluand-specific attitude change and centered engagement scores. There was homoscedasticity, as assessed by visual inspection of a plot of studentized residuals versus unstandardized predicted values. A total of seven outliers were identified as having standardized residuals greater than three standard deviations. However, as no case had a leverage point greater than 0.2, these were retained in the analyses. Similarly, no cases had a Cook's Distance value above 1 (maximum Cook's Distance = 0.59). Examination of P-P Plot indicated approximately normally distributed residual errors.

The addition of the interaction term between engagement scores and evidence strength led to a statistically significant increase in  $R^2$  of 0.015, F(1, 336) = 5.477, p = 0.020. The full model of engagement score, evidence strength, and the interaction term significantly predicted specific attitude change,  $R^2 = 0.089$ , F(3, 336) = 10.986, p < 0.001, adjusted  $R^2 = 0.081$  (See Table 32).

As can be seen in Figure 19, when evidence is strong, individuals tend to demonstrate similar change in specific attitudes regardless of their level of engagement. However, when evidence is weak, those who are not engaged tend to demonstrate negative changes in evaluandspecific attitudes while those who are highly engaged demonstrate positive changes in evaluandspecific attitude.

## Table 32.

Moderation effects of evidence strength on engagement scores predicting specific attitude change.

Step	Variable	r	R <sup>2</sup> Change	В	SE <sub>B</sub>	β
1	Engagement Score	0.145**	0.021**	0.134***	0.039	0.246
2	<b>Evidence Strength</b>	0.239***	0.053***	0.300***	0.067	0.232
	(Weak = 0;					
	Strong=1)					
3	Interaction Term		0.015*	-0.133*	0.057	-0.167
	(Engagement Score					
	x Evidence					
	Strength)					
(Cons	stant)			-0.089	0.048	
*n<0(	$5 \cdot * * n < 01 \cdot n < 0.00$	1				

\*p < 0.05; \*\*p < -.01; p < 0.001

# Figure 19.

Predicted evaluand-specific attitude change based on centered engagement scores and evidence strength.



*Note.* Figure utilizes engagement scores  $\pm 1$  standard deviation from the mean for illustrative purposes. Figure is not a complete description of the data.

Attitude Resilience Analyses. A three-way MANOVA was conducted to assess differences in attitude change and attitude strength as defined by behavioral intent and attitude resilience based on motivation to elaborate, evidence strength, and data visualization approach as well as their interactions. Initial three-way MANOVA model also included donation amount as a dependent variable; however, due to both theoretical differentiation between attitudes and behaviors as well as examination of scatterplots revealed a curvilinear relationship between donation amount and attitude measures, donation amount was removed and analyzed separately. See Table 33 for a summary of descriptive statistics for the dependent variables.

#### Table 33.

Descriptive statistics for dependent variables for attitude resilience three-way MANOVA.

Variable	N	Mean	SD	Min	Max	Skew	Kurtosis
Global Attitude Change	150	-0.07	0.66	-2.80	2.20	-1.11	4.84
Specific Attitude Change	150	0.02	0.70	-3.20	1.60	-1.53	5.35
Behavioral Intent	150	5.67	1.32	1.00	7.00	-1.83	3.25
Specific Attitude Resilience	150	-0.26	0.63	-2.40	2.40	-0.31	3.08

Motivation to elaborate was calculated by averaging three motivation to elaborate items and using the 40<sup>th</sup> and 60<sup>th</sup> percentile cut-offs used to define low and high motivation to elaborate (See Table 34).

#### Table 34.

Descriptive statistics for motivation to elaborate variable.

N	Mean	SD	Min	Max	Skew	Kurtosis	40 <sup>th</sup>	60 <sup>th</sup>
150	5.61	1.26	1.00	7.00	-1.219	1.611	5.67	6.00

Once donation amount was removed, examination of scatterplots indicated there was a linear relationship between the dependent variables. There was no evidence of multicollinearity,

as assessed by Pearson correlation (|r| < 0.9). There were 46 univariate outliers identified through examination of boxplot with values greater than 1.5 box-lengths from the edge of the box. However, as these were neither measurement nor entry errors, these were retained in the analyses. There were no multivariate outliers in the data, as assessed by Mahalanobis distance (critical value = 18.47, p > .001). Multivariate normality was assessed using Shapiro-Wilk test using a Bonferroni correction (p < 0.006). A total of 11 instances of broken multivariate normality assumption was detected (See Table 35). However, as the three-way MANOVA is robust to Type I error, analyses proceeded without transformation to the dependent variables (Bray & Maxwell, 1985; Weinfurt, 1995). There was a sufficient sample size for three-way MANOVA analysis, as there were more than 4 individuals per cell (See Table 36).

#### Table 35.

Condition

1	nstances	of bro	oken mul	ltivariate	normality	assumption	assessed	by Sh	aprio-Wil	lk statistic	: with
1	Bonferron	i corr	ection (	p<0.006)							

	Condition					
Motivation	Visualization	Evidence		Shapri		
to Elaborate	Approach	Strength	Dependent Variable	o-Wilk	Df	Sig.
Low	Minimalist	Weak	Specific Attitude Change	0.848	22	0.003
Low	Embellished	Weak	Specific Attitude Change	0.795	21	0.001
Low	Embellished	Weak	Behavioral Intent	0.813	21	0.001
Low	Embellished	Weak	Resilience	0.848	21	0.004
High	Minimalist	Weak	Specific Attitude Change	0.809	15	0.005
High	Minimalist	Weak	Resilience	0.601	15	< 0.001
High	Minimalist	Strong	Specific Attitude Change	0.785	20	0.001
High	Minimalist	Strong	Resilience	0.851	20	0.006
High	Embellished	Weak	Global Attitude Change	0.663	17	< 0.001
High	Embellished	Strong	Specific Attitude Change	0.780	24	< 0.001
High	Embellished	Strong	Resilience	0.640	24	< 0.001

	Motivation to	Strength	
Chart Type	Elaborate	Туре	N
	Low	Weak	22
Minimalist	Low	Strong	
	High	Weak	15
	mgn	Strong	20
	Low	Weak	21
Embellished	LOW	Strong	14
	High	Weak	17
	111211	Strong	24

Final sample sizes for resilience three-way MANOVA by condition.

Table 36.

Homogeneity of covariance matrices assumption was broken, as assessed by Box's M test (Box's M = 275.70, F(70, 22382.98) = 3.555, p < 0.001). As a result, multivariate analyses were determined utilizing Pillai's criterion as it is more robust to unequal covariance matrices (Olson, 1976). The assumption of homogeneity of variance was broken for global attitude change [F(7, 142) = 4.647, p < 0.001], evaluand-specific attitude change [F(7, 142) = 3.131, p = 0.004], and behavioral intent [F(7, 142) = 14.256, p < 0.001]. As a result, significance level was adjusted to 0.01.

No significant three-way interaction was found between data visualization approach, level of motivation to elaborate, and evidence strength (Pillai's Trace = 0.012, F(4, 139) = 0.417, p = 0.796, partial  $\eta^2 = 0.012$ ). There were also no significant two-way interactions found (See Table 37).

Table 37.

Results of multivariate tests for two-way interactions.

Interaction	Pillai's	F-		Partial
	Trace	value	Sig.	$\eta^2$
Motivation to elaborate x Evidence Strength	0.054	1.968	0.103	0.054
Data Visualization Approach x Evidence Strength	0.048	1.748	0.143	0.048
Data Visualization Approach x Motivation to Elaborate	0.028	1.002	0.409	0.028

Multivariate tests revealed a significant main effect of level of motivation to elaborate (Pillai's Trace = 0.319, F(4, 139) = 16.253, p < 0.001, partial  $\eta^2 = 0.319$ ) and evidence strength (Pillai's Trace = 0.128, F(4, 139) = 5.088, p = 0.001, partial  $\eta^2 = 0.128$ ), but not data visualization approach (Pillai's Trace = 0.029, F(4, 139) = 1.034, p = 0.392, partial  $\eta^2 = 0.028$ ). Follow-up tests of between-subjects effects revealed statistically significant differences in behavioral intent scores based on level of motivation to elaborate (F(1, 142) = 48.752, p < 0.001, partial  $\eta^2 = 0.256$ ). Those who were low in motivation to elaborate reported significantly lower behavioral intent to support the evaluand than those who were high in motivation to elaborate (See Figure 20). The remainder of the outcome variables did not differ significantly based on level of motivation to elaborate (p < 0.01).

## Figure 20.

Average behavioral intent scores based on level of motivation to elaborate.



Follow-up tests of between-subjects effects revealed statistically significant differences in global attitude change based on level of evidence strength (F(1, 142) = 3.101, p = 0.008, partial  $\eta^2 = 0.048$ ). Consistent with regression findings, those who were presented weak evidence tended to lower global attitudes while those who were presented strong evidence tended to increase their global attitudes (See Figure 21). Analysis also indicated that evidence strength had a significant effect on evaluand-specific attitude change, F(1, 142) = 8.193, p < 0.001, partial  $\eta^2 = 0.119$ ; However, given the significant interaction between engagement scores and evidence strength on evaluand-specific attitude change found through the hierarchical multiple regression, this main effect is dismissed. The remainder of the outcome variables did not differ significantly based on level of motivation to elaborate (p < 0.01).

## Figure 21.





Finally, given Phase 1 findings that minimalist and embellished data visualization approaches did not differ in terms of participants' engagement with data visualization, an exploratory analysis was conducted to examine the relationship between elaboration (measured through an interaction between engagement scores and level of evidence strength) and attitude resilience through hierarchical multiple regression. The resilience subsample participants overall had moderately high engagement scores (See Table 38).

#### Table 38.

Descriptive statistics for resilience subsample data visualization engagement scores.

Outcome	N	Mean	SD	Min	Max	Skew	Kurtosis
Engagement	150	5.50	1.24	2.05	8.00	-0.32	-0.55

There was independence of residuals (Durbin-Watson = 2.045). Examination of studentized residual plots and partial regression plots confirmed linear relationships between attitude resilience and centered engagement scores. There was homoscedasticity, as assessed by visual inspection of a plot of studentized residuals versus unstandardized predicted values.

A total of two outliers were identified as having standardized residuals greater than three standard deviations. However, as no case had a leverage point greater than 0.2, these were retained in the analyses. Similarly, no cases had a Cook's Distance value above 1 (maximum Cook's Distance = 0.15). Examination of P-P Plot indicated approximately normally distributed residual errors. The addition of the interaction term between engagement scores and evidence strength did not lead to a statistically significant increase in  $R^2$ , F(1, 146) = 0.00, p= 0.989. The full model of engagement score, evidence strength, and the interaction term did not significantly predict attitude resilience,  $R^2 = 0.018$ , F(3, 146) = 0.888, p = 0.449, adjusted  $R^2 = -0.002$  (See Table 39).

## Table 39.

0.056 0.128		R <sup>2</sup> Change	r	Variable	Step
0.050 0.120	0.066	0.014	0.117	Engagement Score	1
0.105 -0.066	-0.084	0.004	-0.046	Evidence Strength	2
				(Weak $= 0$ ;	
				Strong=1)	
0.086 -0.001	-0.001	0.000		Interaction Term	3
				(Engagement Score	
				x Evidence	
				Strength)	
0.074	-0.214			stant)	(Cons
0.074	-0.214			x Evidence Strength) stant)	(Cons

Moderation effects of evidence strength on engagement scores predicting attitude resilience.

\*p < 0.05; \*\*p < -.01; p < 0.001; Cumulative  $R^2 = 0.018$ ; Adjusted  $R^2 = -0.002$ ; Engagement is 'centered' to a mean of 5.50.

Attitude Persistence Analyses. Out of the 211 participants selected for follow-up, 132 participants (62.5% retention) completed the attitude persistence section of the study by responding to a follow-up survey after a week delay. A three-way MANOVA was conducted to assess differences in attitude change and attitude strength as defined by behavioral intent and attitude persistence based on motivation to elaborate, evidence strength, and data visualization approach as well as their interactions. See Table 40 for a summary of descriptive statistics for the dependent variables.

## Table 40.

Descriptive statistics for dependent variables for attitude persistence three-way MANOVA.

Variable	N	Mean	SD	Min	Max	Skew	Kurtosis
Global Attitude Change	211	0.10	0.63	-2.00	5.20	3.09	22.88
Evaluand-specific Attitude Change	211	0.09	0.59	-2.60	2.40	-0.14	4.92
Behavioral Intent	211	5.55	1.17	1.00	7.00	-1.26	1.54
Global Attitude Persistence	132	0.07	0.55	-1.80	2.20	0.43	2.15
Evaluand-specific Attitude Persistence	132	-0.01	0.60	-1.60	2.80	1.15	4.51

Motivation to elaborate was calculated by averaging three motivation to elaborate items and using the 40<sup>th</sup> and 60<sup>th</sup> percentile cut-offs used to define low and high motivation to elaborate (See Table 41).

#### Table 41.

Descriptive statistics for motivation to elaborate variable.

N	Mean	SD	Min	Max	Skew	Kurtosis	40 <sup>th</sup>	60 <sup>th</sup>
211	5.45	1.27	1.00	7.00	0.17	1.68	5.33	6.00

At first, donation amount was included in the three-way MANOVA analyses; however, examination of scatterplots revealed that it had a curvilinear relationship with several other dependent variables and was therefore removed from the MANOVA analyses. Once donation amount was removed, examination of scatterplots indicated there was a linear relationship between the dependent variables. Additionally, multicollinearity was initially detected using Pearson correlation between global and evaluand-specific attitude persistence scores for the low motivation to elaborate group who were presented minimalist visualizations with weak evidence, |r| = 0.917.

There were 33 univariate outliers identified through examination of boxplot with values greater than 1.5 box-lengths from the edge of the box. However, as these were neither measurement nor entry errors, these were retained in the analyses. There were no multivariate outliers in the data, as assessed by Mahalanobis distance (critical value = 20.52, p > .001). Multivariate normality was assessed using Shapiro-Wilk test using a Bonferroni correction (p < 0.006). A total of five instances of broken multivariate normality assumption was detected (See Table 42). However, as the three-way MANOVA is robust to Type I error, analyses proceeded without transformation to the dependent variables (Bray & Maxwell, 1985; Weinfurt, 1995).

There was a sufficient sample size for three-way MANOVA analysis, as there were more than five individuals per cell (See Table 43).

Homogeneity of covariance matrices assumption was broken, as assessed by Box's M test (Box's M = 310.80, F(105, 6308.58) = 2.31, p < 0.001). As a result, multivariate analyses were determined utilizing Pillai's criterion as it is more robust to unequal covariance matrices (Olson, 1976). The assumption of homogeneity of variance was broken for global attitude change [F(7, 84) = 3.69, p = 0.002], behavioral intent [F(7, 84) = 5.34, p < 0.001], and evaluand-specific attitude persistence [F(7, 84) = 2.28, p = 0.036]. As a result, significance level was adjusted to 0.01.

## Table 42.

Instances of broken multivariate normality assumption assessed by Shaprio-Wilk statistic with Bonferroni correction (p<0.006).

	Condition					
Motivation	Visualization	Evidence		Shaprio-		
to Elaborate	Approach	Strength	Dependent Variable	Wilk	Df	Sig.
High	Minimalist	Weak	Specific Attitude Change	0.484	16	0.002
High	Minimalist	High	Global Attitude Change	0.521	11	< 0.001
Low	Embellished	Strong	Specific Attitude	0.718	9	0.002
			Persistence			
High	Embellished	Strong	Specific Attitude Change	0.496	16	< 0.001
High	Embellished	Strong	Global Attitude	0.671	16	< 0.001
			Persistence			

## Table 43.

Chart Type	Motivation to Elaborate	Strength Type	N
	Low	Weak	10
Minimalist	Low	Strong	8
winninalist	High	Weak	16
	Ingn	Strong	11
	Low	Weak	10
Embellished	Low	Strong	9
Lindenished	High	Weak	12
		Strong	16

Final sample size for persistence three-way MANOVA.

No significant three-way interaction was found between data visualization approach, level of motivation to elaborate, and evidence strength (Pillai's Trace = 0.010, F(5, 80) = 0.159, p = 0.977, partial  $\eta^2 = 0.010$ ). Additionally, no significant two-way interactions were found (See Table 44).

## Table 44.

Results of multivariate tests for two-way interactions.

Interaction	Pillai's	F-		Partial
	Trace	value	Sig.	$\eta^2$
Motivation to elaborate x Evidence Strength	0.03	0.435	0.823	0.026
Data Visualization Approach x Evidence Strength	0.03	0.431	0.826	0.026
Data Visualization Approach x Motivation to Elaborate	0.02	0.308	0.907	0.019

Multivariate tests revealed a significant main effect of level of motivation to elaborate (Pillai's Trace = 0.567, F(5, 80) = 20.97, p < 0.001, partial  $\eta^2 = 0.567$ ), but not data visualization approach (Pillai's Trace = 0.098, F(5, 80) = 1.74, p = 0.134, partial  $\eta^2 = 0.098$ ) nor evidence strength (Pillai's Trace = 0.103, F(5, 80) = 1.83, p = 0.116, partial  $\eta^2 = 0.103$ ).

Follow-up tests of between-subjects effects revealed statistically significant differences in global attitude change scores and behavioral intent scores based on level of motivation to elaborate (F(1, 84) = 1.422, p = 0.010, partial  $\eta^2 = 0.079$  and F(1, 84) = 63.519, p < 0.001, partial  $\eta^2 = 0.508$ , respectively). The remainder of the outcome variables did not differ significantly based on level of motivation to elaborate (p < 0.01).

Those who were low in motivation to elaborate reported significantly lower behavioral intent to support the evaluand than those who were high in motivation to elaborate (See Figure 22).

#### Figure 22.





In converse to the resilience subsample, those in the persistence subsample who were low in motivation to elaborate tended to increase their global attitudes while those who were high in motivation to elaborated tended to decrease their global attitudes after seeing the study results (See Figure 23).

#### Figure 23.



Mean Global Attitude Change score by motivation to elaborate level.

As before, given Phase 1 findings that minimalist and embellished data visualization approaches did not differ in terms of participants' engagement with data visualization, an exploratory analysis was conducted to examine the relationship between elaboration (measured through an interaction between engagement scores and level of evidence strength) and evaluandspecific attitude persistence through hierarchical multiple regression. The analyses focus solely on evaluand-specific attitude persistence due to Phase 1 findings comparing elaboration between global and evaluand-specific attitudes.

The persistence subsample had moderately high data visualization engagement scores (See Table 45). There was independence of residuals (Durbin-Watson = 1.419). Examination of studentized residual plots and partial regression plots confirmed linear relationships between attitude persistence and centered engagement scores. There was homoscedasticity, as assessed by visual inspection of a plot of studentized residuals versus unstandardized predicted values. Engagement scores were centered to reduce multicollinearity. Pearson's correlations indicated no multicollinearity (Pearson's |r|<0.7) and the minimum Tolerance statistic was 0.521.

Outcome	N	Mean	SD	Min	Max	Skew	Kurtosis
Engagement	132	5.37	1.11	2.55	8.00	-0.25	-0.38

**Table 45.**Descriptive statistics of engagement scores for persistence subsample.

A total of two outliers were identified as having standardized residuals greater than the standard deviations. However, as no case had a leverage point greater than 0.2, these were retained in the analyses. Similarly, no cases had a Cook's Distance value above 1 (maximum Cook's Distance = 0.09). Examination of P-P Plot indicated approximately normally distributed residual errors.

The addition of the interaction term between engagement scores and evidence strength did not lead to a statistically significant increase in  $R^2$ , F(1, 128) = 0.17, p = 0.130. The full model of engagement score, evidence strength, and the interaction term did not significantly predict evaluand-specific attitude persistence,  $R^2 = 0.050$ , F(3, 128) = 2.240, p = 0.087, adjusted  $R^2 = 0.028$  (See Table 46).

## Table 46.

Moderation effects of evidence strength on engagement scores predicting attitude persistence.

Step	Variable	r	R <sup>2</sup> Change	В	SE <sub>B</sub>	β		
1	Engagement Score	-0.056	0.003	0.0.047	0.065	0.087		
2	<b>Evidence Strength</b>	-0.176*	0.030	-0.207	0.104	-0.172		
	(Weak $= 0$ ;							
	Strong=1)							
3	Interaction Term		0.017	-0.143	0.094	-0.182		
	(Engagement Score							
	x Evidence							
	Strength)							
(Cons	stant)			0.107	0.074			
*n < 0 (	$p < 0.05$ , ** $p < 0.01$ , $p < 0.001$ , Cumulativa $P^2 = 0.050$ ; A divised $P^2 = 0.029$ ; Engagement is							

\*p < 0.05; \*\*p < -.01; p < 0.001; Cumulative  $R^2 = 0.050$ ; Adjusted  $R^2 = 0.028$ ; Engagement is 'centered' to a mean of 5.37.

**Behavioral Intent Analyses.** In addition to examining the relationship between behavioral intent scores and motivation to elaborate, evidence strength, and data visualization approach through

the previous MANOVA analyses, the relationship between behavioral intent and elaboration was examined through exploratory hierarchical multiple regression analyses. Based on hypothesized evaluation influence pathways, elaboration was examined using an interaction term between data visualization engagement and evidence strength on behavioral intent scores. Table 47 presents engagement scores for the overall sample.

#### Table 47.

Descriptive statistics of engagement scores for overall Phase 2 sample.

Outcome	N	Mean	SD	Min	Max	Skew	Kurtosis
Engagement	340	5.32	1.19	2.05	8.00	-0.12	-0.50

There was independence of residuals (Durbin-Watson = 1.911). Examination of studentized residual plots and partial regression plots confirmed linear relationships between behavioral intent scores and centered engagement scores. There was homoscedasticity, as assessed by visual inspection of a plot of studentized residuals versus unstandardized predicted values. Engagement scores were centered to reduce multicollinearity. Pearson's correlations indicated no multicollinearity (Pearson's |r|<0.7) and the minimum Tolerance statistic was 0.529.

A total of five outliers were identified as having standardized residuals greater than three standard deviations. However, as no case had a leverage point greater than 0.2, these were retained in the analyses. Similarly, no cases had a Cook's Distance value above 1 (maximum Cook's Distance = 0.08). Examination of P-P Plot indicated approximately normally distributed residual errors.

The addition of the interaction term between engagement scores and evidence strength led to a statistically significant increase in  $R^2$  of 0.010, F(1, 336) = 5.851, p = 0.016. The full

model of engagement score, evidence strength, and the interaction term significantly predicted

behavioral intent,  $R^2 = 0.420$ , F(3, 336) = 81.002, p < 0.001, adjusted  $R^2 = 0.415$  (See Table 48).

## Table 48.

Moderation effects of evidence strength on engagement scores predicting behavioral intent to support the evaluand.

Step	Variable	r	R <sup>2</sup> Change	В	SE <sub>B</sub>	β
1	Engagement Score	0.638***	0.408	0.773***	0.060	0.730***
2	<b>Evidence Strength</b>	0.083	0.002	0.112	0.105	0.045
	(Weak $= 0$ ;					
	Strong=1)					
3	Interaction Term	0.365***	0.010	-0.214*	0.088	-0.138*
	(Engagement Score					
	x Evidence					
	Strength)					
(Constant)				5.512	0.074	

\*p < 0.05; \*\*p < -.01; p < 0.001; Cumulative  $R^2 = 0.420$ ; Adjusted  $R^2 = 0.415$ ; Engagement is 'centered' to a mean of 5.32.

Based on results, behavioral intention is predicted based on the following regression

equation:

# **Equation 4.**

Regression equation predicting behavioral intent to support the evaluand.

 $\widehat{Y} = 5.512 + 0.773$ (Engagement Score) + 0.112 (Evidence Strength Level)

- 0.214(Engagement Score x Evidence Strength Leve)

As demonstrated in Figure 24 below, individuals who find the visualization highly

engaging show less intent to support the evaluand when evidence is strong compared to when evidence is weak. The reverse is true of those who do not find the visualization engaging; these individuals show greater intent to support the program when evaluation findings provide strong evidence compared to weak evidence.

Figure 24.



Predicted behavioral intent based on centered engagement scores and evidence strength.

**Note.** Figure utilizes engagement scores  $\pm 1$  standard deviation from the mean for illustrative purposes. Figure is not a complete description of the data.

**Donation Behavior Analyses.** All 340 participants were given the opportunity to respond whether they would like to donate and, if so, the amount they would like to donate to the WISH program. Of these, 165 (48.5%) decided to donate to the program. Table 49 summarizes descriptive statistics for donation amount. A three-way ANOVA was conducted to assess the impact of motivation to elaborate, data visualization approach, and evidence strength on donation amounts.

## Table 49.

Descriptive statistics for donation amount.

Variable	N	Mean	SD	Min	Max	Skew	Kurtosis
Donation Amount	340	\$0.67	\$0.88	\$0.00	\$2.40	0.97	-0.65
As before, motivation to elaborate was calculated by averaging three motivation to elaborate items and utilizing the 40th and 60th percentile cut-offs used to define low and high motivation to elaborate (See Table 50).

#### Table 50.

Descriptive statistics for motivation to elaborate scores for donation behavior analyses.

N	Mean	SD	Min	Max	Skew	Kurtosis	40 <sup>th</sup>	60 <sup>th</sup>
340	5.47	1.28	1.00	7.00	-1.19	1.44	5.33	6.00

There were 18 univariate outliers identified through examination of boxplot with values

greater than 1.5 box-lengths from the edge of the box. However, as these were neither

measurement nor entry errors, these were retained in the analyses. Normality of dependent

variable distribution was assessed using Shapiro-Wilk test using a Bonferroni correction (p < p

0.006). In all instances, the normal distribution assumption was violated (See Table 51).

Similarly, the assumption of homogeneity of variance was also broken (Levene's F(7, 249) =

12.82, *p* < 0.001).

a 1...

#### Table 51.

Instances of broken multivariate normality assumption assessed by Shaprio-Wilk statistic with Bonferroni correction (p < 0.006).

	Condition					
Motivation	Visualization	Evidence	Shaprio-	Df	Sig	
to Elaborate	Approach	Strength	Wilk	Dj	515.	
Low	Minimalist	Weak	0.440	28	< 0.001	
High	Minimalist	Weak	0.773	39	< 0.001	
Low	Embellished	Weak	0.390	25	< 0.001	
High	Embellished	Weak	0.803	36	< 0.001	
Low	Minimalist	Strong	0.530	26	< 0.001	
High	Minimalist	Strong	0.805	37	< 0.001	
Low	Embellished	Strong	0.575	25	< 0.001	
High	Embellished	Strong	0.837	41	< 0.001	

A logarithmic transformation was conducted to the donation amount variable. However, results of the three-way ANOVA were the same regardless of whether the original or transformed variable was utilized, finding no significant three-way or two-way interactions, and finding only a significant main effect of motivation to elaborate (See Table 52 for comparison of results). Follow-up independent samples t-test analyses indicate that those who were low in motivation to elaborate donated on average \$0.68 less than those who had high motivation to elaborate [CI95%(-0.88 to -0.46), t(253.11) = -7.02, p < 0.001].

#### Table 52.

Comparison of three-way ANOVA findings using original and transformed donation amount.

	Origir	al Donation	Amount	Log	10 Transfo nation Am	rmed ount
	<i>F</i> -value	Sig.	Partial $\eta^2$	<i>F</i> -value	Sig.	Partial $\eta^2$
Three-way interaction: Motivation to Elaborate, Data Visualization Approach, and Evidence Strength	0.136	0.713	0.001	0.339	0.561	0.003
Two-way interaction: Motivation to Elaborate and Data Visualization Approach	0.272	0.603	0.001	0.376	0.541	0.003
Two-way interaction: Motivation to Elaborate and Evidence Strength	0.340	0.560	0.001	0.241	0.625	0.002
Two-way interaction: Data Visualization Approach and Evidence Strength	0.981	0.323	0.004	0.179	0.673	0.002
Main Effect: Data Visualization Approach	0.111	0.739	0.000	0.177	0.675	0.002
Main Effect: Evidence Strength	1.183	0.278	0.005	0.436	0.510	0.004
Main Effect: Motivation to Elaborate	40.465	<0.001***	0.140	7.581	0.007**	0.062

Note. \*\**p*<0.01. \*\*\**p*<0.001

**Influence Pathway Analyses.** In addition, the predicted relationship between post-intervention attitudes and donation behavior was explored through multiple regression analyses. Summary statistics for post intervention global and evaluand-specific attitudes are included in Table 53.

# Table 53.

Variable	N	Mean	SD	Min	Max	Skew	Kurtosis
Post Intervention	257	6.14	1.01	1.00	7.00	-1.81	4 21
Global Attitude	231	0.14	1.01	1.00	7.00	-1.01	7.21
Post Intervention							
Evaluand-specific	257	5.57	1.26	2.40	7.00	-0.36	-1.09
Attitude							

Descriptive Statistics for post intervention attitudes.

There was independence of residuals (Durbin-Watson = 0.822) and examination of scatterplots revealed linear relationships between donation amount and the independent variables. There was homoscedasticity, as assessed by visual inspection of a plot of studentized residuals versus unstandardized predicted values. No multicollinearity was detected, as assessed by Tolerance values less than 0.1 (minimum tolerance value was 0.610). One outlier was identified based on examination of standardized residuals greater than  $\pm 3$  standard deviations. The leverage value of this outlier was less than 0.007, far less than the cutoff value of 0.2, and the case was retained in the analysis. There were no cases with a leverage value greater than 0.2 nor a Cook's Distance value of greater than 1 (Cook & Weisberg, 1982). Finally, examination of Normal P-P Plot of standardized residuals indicated normal distribution.

Results of the multiple regression analyses indicated that both global and evaluandspecific post-intervention attitudes significantly predicted donation amounts, F(2, 254) = 53.529, p < 0.001, adjusted  $R^2 = 0.291$  (See Table 54).

# Table 54.

Donation Amount	В	95% CI for <i>B</i>		SE B	β	$R^2$	$\Delta R^2$
		LL	UL				
Model						0.297	0.291
Constant	0.430			0.289			
Global Post-intervention Attitude	0.473***	0.357	0.589	0.059	0.541***		
Evaluand-Specific Post- intervention Attitude	-0.482***	-0.576	-0.388	0.048	-0.682***		

Effects of post-intervention attitudes on predicting donation amount.

*Note*. Model = "Enter" method in SPSS Statistics; B = unstandardized regression coefficient; CI = confidence interval; LL = lower limit; UL = upper limit; *SE B* = standard error of the coefficient;  $\beta$  = standardized coefficient;  $R^2$  = coefficient of determination;  $\Delta R^2$  = adjusted  $R^2$ . \*p < 0.05. \*\*p < 0.01. \*\*\*p < 0.001.

Figure 25 and Figure 26 illustrate the effects of post-intervention attitudes on donation

behavior utilizing the following regression equation:

# **Equation 5.**

Regression equation predicting donation amount.

 $\widehat{Y} = 0.430 + 0.473$ (Global Post – intervention attitude)

- 0.482(evaluand - specific Post - intervention attitude)

# Figure 25.

Effect of global post-intervention attitudes on donation behavior



Note. Figure uses  $\pm 1$  standard deviation for post-interventional global attitude scores and holds constant evaluand-specific post-intervention attitudes at the mean for illustrative purposes. Figure is not a complete description of the data.

#### Figure 26.

Effect of evaluand-specific post-intervention attitudes on donation behavior.



Note. Figure uses  $\pm 1$  standard deviation for post-interventional evaluand-specific attitude scores and holds constant post-intervention global attitudes at the mean for illustrative purposes. Figure is not a complete description of the data.

## DISCUSSION

Phase 2 explores the impact of motivation to elaborate, evidence strength, and data visualization approaches on attitude change, attitude strength, and behaviors. Additionally, Phase 2 tests the hypothesized pathway from attitudes to behaviors. Overall, Phase 2 findings provide support for the hypothesized pathway from attitudes to behaviors, but do not extend Phase 1 findings linking elaboration to attitude strength. Additionally, Phase 2 findings continue to emphasize the role of motivation to elaborate on key outcomes; here, motivation to elaborate played a role in behavioral intent and donation behaviors.

While Phase 1 findings established that engagement connects to elaboration and attitude change, Phase 2 replicated these findings only for evaluand-specific attitude change. Additionally, Phase 2 found partial evidence that this relationship extends to attitude strength. No significant differences in attitude resilience and persistence were found based on the interaction of data visualization engagement and evidence strength. Only behavioral intent, the final component of attitude strength, was predicted by elaboration. These findings are surprising considering previous research which indicates that stronger attitudes are formed through central route elaboration processes (e.g., Crano, 1995; Haugtvedt, Schumann, Schneier, and Warren, 1994; Petty & Cacioppo, 1986; Verplanken, 1991). Indeed, none of the explored factors impacted attitude resilience and persistence, and further research is needed to understand how attitude strength is formed within an evaluation context. The implications may be that evaluation influence pathways which take more time may be less likely to form. As a result, it may be beneficial to time presentation of findings as close as possible to opportunities for stakeholders to act on the findings. Alternatively, it may be possible that other aspects of an evaluation may facilitate attitude persistence and resilience, while this study was limited to only a brief written overview of evaluation findings.

Additionally, Phase 2 findings built on those from Phase 1 by adding the final step of the hypothesized evaluation influence pathway – behaviors. Attitudes measured after participants had seen the evaluation findings significantly predicted donation behaviors, in alignment with the Theory of Changed Behavior (Ajzen, 1991) and providing empirical evidence for evaluation influence pathways outlined in Mark and Henry's (2004) Evaluation Influence framework.

Phase 2 findings also continued to emphasize a key role for motivation to elaborate in behavioral outcomes. Those who were high in motivation to elaborate expressed greater intent to support the evaluand and donated more money than those who were low in motivation to elaborate. The Theory of Changed Behavior notes that in addition to attitudes, behaviors are guided by social norms and perceived behavioral control (Ajzen, 1991). Motivation to elaborate was defined by the level of relevance the hypothetical evaluand had for an individual. It is

131

possible that this characteristic may be reflective of a shared group (e.g., women unsatisfied with work-life balance, etc.) with societal norms that guide behavior.

Finally, Phase 2 findings regarding data visualization replicate and extend those found in Phase 1. Neither minimalist nor embellished data visualizations impacted elaboration or attitude change and were not factors that promoted greater attitude strength or donation behaviors. In combination with Phase 1 findings that these two data visualization approaches did not differ in levels of engagement, neither of these visualization conventions increase involvement in a manner that promotes motivation to elaborate (Petty & Cacioppo, 1986). As a result, Phase 3 of this research explores an alternative approach–interactive data presentations – that directly increases audience involvement.

# CHAPTER 4 PHASE III: INTERACTIVE DATA DISPLAY AFFECT ON ELABORATION, ATTITUDE STRENGTH, AND BEHAVIOR

The final phase of this research was informed by four key findings learned from the initial two phases. The first finding that informed the direction of the third phase of research was that there was no evidence of differences in levels of engagement or elaboration produced by minimalist and embellished data visualizations. This suggests that neither data visualization convention triggers the elaboration mechanism identified in Mark and Henry's (2004) Evaluation Influence framework. However, it is possible that other data presentation approaches may be more effective at promoting engagement and elaboration. Interactivity has been described as a method of turning the audience from "a passive consumer into an active participant" (Weissgerber et al., 2017, p. 20592). Similarly, Natter and Berry (2005) found that communication which incorporates active participation can facilitate understanding of information. Thus, while interactive data displays are not as common in evaluation products, this approach may hold promise as a method to increase motivation to elaborate by increasing participants' personal involvement. To this end, the third phase of this research explores whether an interactive presentation of findings results in greater engagement and elaboration than minimalist data visualization approaches.

A second finding that informed the third phase of this research was that those who were highly motivated to elaborate took longer to examine weak evidence. As discussed in Phase 1, it is possible that these individuals were surprised by the findings given that they also tended to have more positive initial attitudes towards the evaluand, and that this cognitive dissonance prompted a closer examination of the findings. In research by Zanna and Cooper (1974) and Petty, Fabrigar, and Wegener (2003), cognitive dissonance has been noted as a motivator for

133

elaboration and attitude change. As a result, this phase of research explores how prior expectations of evaluation results impact attitudinal and behavioral outcomes.

The third finding was that in the second phase, elaboration did not result in differences in attitude resilience and persistence. This phase of research seeks to determine whether interactive data presentation affects the degree to which individuals who elaborate can resist counter-messaging and maintain their attitudes over time.

Finally, in the second phase, findings indicated that donation behavior was driven largely by characteristics of the participants – specifically, their motivation to elaborate – rather than the strength of the evaluation findings. This phase of research seeks to determine if an interactive presentation of findings results in greater emphasis on the evaluation findings guiding donation behaviors.

In essence, Phase 3 of the research re-examines the hypothesized pathway from data presentation to behavior presented in Figure 4, comparing interactive and minimalist approaches to the presentation of evaluation findings, with the following hypotheses:

- 4. Do lessons learned translate to interactive data displays?
  - a. *Hypothesis 4.1:* Interactive data presentation will be perceived as more engaging than static minimalist data visualizations. Additionally, those individuals with high motivation to elaborate will report higher engagement than those with low motivation to elaborate.
  - b. *Hypothesis 4.2:* Interactive data presentation will result in greater attitude change when evaluation findings are strong, but not when evaluation findings are weak; Static minimalist data visualization will result in similar attitude change regardless of evidence strength.

- c. *Hypothesis 4.3:* The degree to which participants' expectations of program results differ from the evaluation's findings will predict the extent of attitude change.
- d. *Hypothesis 4.4:* Participants shown interactive data presentations with strong evaluation findings will display greater attitude persistence and resilience than those shown weak evidence. Participants shown static minimalist data visualizations will show similar attitude persistence and resilience regardless of the strength of evaluation findings.
- e. *Hypothesis 4.5:* Participants shown interactive data presentations with strong evidence will donate more than those shown weak evaluation findings. Participants shown static minimalist data visualizations will show similar donation behaviors regardless of the strength of evaluation findings.

#### **METHODS**

**Recruitment.** As in the previous phases of this research, the Phase 3 study recruited participants from Amazon Mechanical Turk (MTurk), a crowdsourcing platform. The study participants were limited to Amazon MTurk workers who reside in the United States to minimize potential confounding factors and the risk of language barriers as the study was conducted in American English. Recruitment text is included in Appendix B. Recruitment occurred in February 2020, and participants were paid a \$2.40 incentive upon completion of the core questionnaire, and an additional \$1.60 bonus incentive for additionally completing either the resilience questionnaire or follow-up persistence questionnaire. Incentives were set based on estimated time needed to complete the survey and minimum wage in Fall 2020 in the United States. Participants who were randomized into the resilience condition were immediately presented the resilience questionnaire; if they did not wish to participate in the resilience measure, they were asked to skip these questions without answering to submit their survey

responses. Those in the follow-up persistence condition received an email communication through Amazon MTurk's API one week after completing the initial survey inviting them to complete the follow-up survey (please see Appendix B).

**Design.** As in the previous studies, Phase 3 used a 2 (motivation to elaborate: low vs. high) x 2 (data presentation approach: interactive vs. minimalist) x 2 (evidence strength: weak vs. strong) between-subjects experimental design.

**Procedure.** The procedure replicated that used in Phase 2.

*Motivation to Elaborate*. The approach to measuring motivation to elaborate replicated that used in the previous phases of this research.

*Scenario.* The same scenario was utilized as in previous phases of this research. As in Phase 2, the reveal of the simulated nature of the WISH evaluation study was delayed to capture more authentic donation behaviors. The simulated nature of the scenario was revealed to participants during debrief, as in Phase 2.

*Attitude Measure.* Phase 3 utilized the same attitude measure as in previous phases of this research.

*Data Display.* Phase 3 utilized the same minimalist data visualization as previous phases of this research. However, rather than using embellished data visualization, Phase 3 utilized an interactive question-and-answer approach to presenting the evaluation findings. Fleming (2011) posed that inclusion of thought-provoking questions or probes in reports may be able to support thoughtful processing of information. To explore this, for each of the four findings (WISH participants' change in work-life balance, Comparison group's change in work-life balance, WISH participants' change in median income, Comparison group's change in median income), the interactive data presentation first asked participants to guess the results of the study (e.g.,

136

"Prior to joining the WISH program, 10% of participants reported they were "extremely satisfied" with their work-life balance. What percent do you think reported being "extremely satisfied" with their work life balance after completing the WISH program?"; See Appendix K for interactive data presentation tool). For each guess, participants respond on a 0% to 100% scale. After participants submit each guess, an embellished visual of the evaluation finding was displayed, with accompanying narrative re-iterating their guess and stating the evaluation finding.

*Evidence Strength.* Phase 3 utilized the same effect sizes as in the previous two phases of this research.

*Engagement Measure*. Phase 3 utilized the same engagement measure as the previous two phases of this research, with slight wording changes. Rather than framing questions as "while reading this chart", questions were re-framed more broadly as "while reading this presentation of findings" to be inclusive of the interactive data presentation approach. Final Phase 3 Engagement Measure and scale reliability analyses are included in Appendix H.

*Behavioral Intent Survey*. Phase 3 utilized the same behavioral intent survey as in the second phase of this research.

*Measure of Donation Behavior*. Phase 3 utilized the same measures of donation behavior as in the second phase of this research.

*Demographics.* The same demographics questions were used as in the two previous phases of this research.

*Attitude Persistence.* Phase 3 utilized the same measure of attitude persistence as in the second phase of this research.

137

*Attitude Resilience.* Phase 3 utilized the same measure of attitude resilience as in the second phase of this research.

*Difference Between Expectations and Evaluation Findings.* In research by Zanna and Cooper (1974) and Petty, Fabrigar, and Wegener (2003), cognitive dissonance has been noted as a motivator for elaboration and attitude change. This was measured in two ways in Phase 3: first, to measure perceptions of the findings, a 7-point semantic differential question was added to determine how unsurprising or surprising participants found the WISH evaluation study findings. Second, to explore whether an interactive data presentation such as the one utilized in this study can trigger cognitive dissonance, a difference score was calculated between participants' estimations and the evaluation findings. The difference score was used to predict attitude change.

#### RESULTS

**Population.** As in the previous studies, Phase 3 continued to recruit Amazon MTurk workers. As in the previous phases of this research, the study participants were limited to Amazon MTurk workers who reside in the United States to minimize potential confounding factors and the risk of language barriers as the study was conducted in American English. Based on power analyses, a total participant count of 147 participants was needed to detect a medium effect size (f = 0.3). However, given possible attrition during the follow-up phase of the study and the need to split the sample during the follow-up to minimize research fatigue, the study recruited 436 participants. Of these, 176 were randomly assigned to complete the resilience task while 260 were randomly assigned to complete the persistence task. Of the 260 recruited persistence respondents, 180 completed the follow-up questionnaire after a week delay, resulting in a 69.2% retention rate. There were no significant differences in retention between conditions ( $\chi^2(3) = 2.318, p = 0.509$ ).

A total of eight participants were removed from the sample because their responses to the gender demographic question did not match the gender reported in their Amazon Mturk profile, reducing the overall sample size to 428, the persistence subsample to 172, and the recruited persistence subsample to 256. This section presents the demographics for both the overall sample and each subsample.

As can be seen in Table 55, the overall sample contained roughly equal representation from male and female participants. There were no significant differences in persistence subsample retention based on gender,  $\chi^2(2) = 1.021$ , p = 0.600. Study condition did not differ significantly based on gender for the overall sample ( $\chi^2(3) = 0.091$ , p = 0.993), the resilience subsample ( $\chi^2(3) = 0.117$ , p = 0.990), the persistence recruited subsample ( $\chi^2(3) = 0.221$ , p = 0.974), nor the retained persistence subsample ( $\chi^2(3) = 2.318$ , p = 0.509).

### Table 55.

Gend	ler d	lemograpl	hics f	or P	hase 3	' research	h sampl	le and	l subsamp	les.

Sample	N	Male	Female
Overall Sample	426	215 (50.5%)	211 (49.5%)
Resilience Subsample	172	85 (50.6%)	87 (49.4%)
Persistence Subsample Recruited	256	128 (50.0%)	128 (50.0%)
Persistence Subsample Retained	180	91 (50.6%)	87 (48.3%)
		68.7% retained	68.0% retained

Most respondents in the overall sample were either 25-34 years old (40.1%) or 35-44 years old (28.9%, See Table 56). One participant declined to report their age. Persistence subsample retention did no differ based on age ( $\chi^2(7) = 7.890, p = 0.342$ ). Based on Pearson Chi-Square analyses, there were no significant differences in age group based on study condition in the overall sample ( $\chi^2(21) = 15.996, p = 0.207$ ), resilience subsample ( $\chi^2(18) =$ 

14.082, p = 0.724), recruited persistence subsample ( $\chi^2(18) = 22.220, p = 0.222$ ), nor retained persistence subsample ( $\chi^2(18) = 21.198, p = 0.270$ ).

# Table 56.

Phase 3 participants by age groups.

Age (years)	18 to 24	25 to 34	35 to 44	45 to 54	55 to 64	65 t0 74	75 +
Overall Sample ( $N = 426$ )	4.9%	40.1%	28.9%	14.3%	8.2%	3.1%	0.2%
Resilience Subsample ( $N = 172$ )	5.8%	37.2%	28.5%	12.8%	9.3%	5.8%	0.0%
Persistence Subsample Recruited ( $N = 256$ )	4.3%	42.2%	29.3%	15.2%	7.4%	1.2%	0.4%
Persistence Subsample Retained $(N = 180)$	4.4%	38.9%	32.8%	15.0%	7.2%	1.1%	0.6%

Additionally, most participants in the overall sample (42.7%) had completed a bachelor's degree and an additional 12.9% had completed a master's degree or professional degree (See Table 57). Similar trends held for the subsamples. There was no significant difference in persistence subsample retention based on educational level ( $\chi^2(6) = 9.623, p = 0.141$ ). Based on Pearson Chi-Squared analyses, there were no significant differences in education level across conditions in the overall sample ( $\chi^2(18) = 14.671, p = 0.684$ ), resilience subsample ( $\chi^2(15) = 7.493, p = 0.943$ ), recruited persistence subsample ( $\chi^2(15) = 14.407, p = 0.495$ ).

#### Table 57.

			Persistence	Persistence
	Overall	Resilience	Subsample	Subsample
	Sample	Subsample	Recruited	Retained
N	426	172	256	180
Some High School	0.0%	0.0%	0.0%	0.0%
High School Diploma/	16 20/	16.00/	15 60/	15 00/
GED	10.2%	10.9%	15.0%	15.0%
Associate's Degree/	27.00/	24 40/	28 50/	22.00/
Some College	27.0%	24.4%	28.3%	25.9%
Bachelor's Degree	42.7%	45.3%	41.0%	45.6%
Master's Degree/	12 00/	11 60/	14 10/	14 40/
Professional Degree	12.9%	11.0%	14.1%	14.4%
Doctorate/PhD	0.7%	1.2%	0.4%	0.6%

Phase 3 participants' highest level of competed education.

Demographic questions also explored participants' familiarity with data visualization through three questions utilizing a six-point frequency scale. Overall, results indicate that participants tended to see visualized data (M = 4.00, SD = 1.32) and use data displayed in charts to make decisions (M = 3.38, SD = 1.43) on a weekly basis but did not typically create data visualizations themselves (M = 2.59, SD = 1.47). In the persistence subsample, there were significant differences in mean responses to these questions between those that completed the follow-up survey and those that did not (t(258) = -2.002, p = 0.046), such that those that were retained in the study had more experience with data visualization than those that did not (M =3.41, SD = 1.17; M = 3.09, SD = 1.26, respectively). One-way ANOVA analyses revealed no differences in data visualization experience between study conditions in the overall sample (F(3,422) = 0.315, p = 0.815), the resilience subsample (F(3, 168) = 0.352, p = 0.788), the recruited persistence subsample (F(3, 252) = 0.452, p = 0.818), nor the retained persistence subsample (F(3, 176) = 1.949, p = 0.232).

**Engagement Analysis.** All 434 participants reported their level of engagement with the presentation of evaluation findings through 20 engagement items on a (1-strongly disagree to 7-strongly agree) scale, which were averaged to produce an overall engagement score (see Table 58). A three-way ANOVA was conducted in order to assess the effects of data presentation approach, motivation to elaborate, evidence strength as well as their interactions on engagement scores.

#### Table 58.

Phase 3 Participants' Descriptive Statistics of Engagement Scores.

Variable	N	Mean	SD	Min	Max	Skew	Kurtosis
Engagement	434	4.90	0.91	1.75	7.00	-0.64	0.53

In order to assess level of motivation to elaborate, three motivation to elaborate items were averaged and the 40<sup>th</sup> and 60<sup>th</sup> percentile cut-offs used to define low and high motivation to elaborate (see Table 59). A total of 45 individuals had scores that fell between the 40<sup>th</sup> and 60<sup>th</sup> percentiles and were not included in the analyses. Frequencies of categorical variables are presented in Table 60.

# Table 59.

Phase 3 Participants' descriptive statistics of motivation to elaborate scores.

Variable	N	Mean	SD	Min	Max	Skew	Kurtosis	<b>40</b> <sup>th</sup>	60 <sup>th</sup>
Motivation to	434	5.03	1.51	1.00	7.00	-0.85	0.11	5.00	5.67
Elaborate									

There was independence of observations by study design. There were 10 univariate outliers identified through examination of boxplot with values greater than 1.5 box-lengths from

the edge of the box. However, as these were neither measurement nor entry errors, these were retained in the analyses.

# Table 60.

Frequency of categorical independent variables in three-way ANOVA assessing engagement.

	Motiva	ation to	Data Pre	sentation	Evidanca Strangth		
	Elaborate		Арри	roach	Evidence Strength		
	Low	High	Minimalist	Interactive	Weak	Strong	
N	198	191	192	197	204	185	

Normality of dependent variable distribution was assessed using Shapiro-Wilk test using a Bonferroni correction (p < 0.006). The normal distribution assumption was violated once (See Table 61). The assumption of homogeneity of variance was met (Levene's F(7, 381) = 0.809, p = 0.580).

#### 0.500).

# Table 61.

Instances of broken multivariate normality assumption assessed by Shaprio-Wilk statistic with Bonferroni correction (p<0.006).

	Condition				
Motivation	Presentation	Evidence	Shaprio-		
to Elaborate	Approach	Strength	Wilk	Df	Sig.
Low	Interactive	Weak	0.911	56	0.001

There was no significant three-way interaction of data presentation approach, motivation to elaborate, and evidence strength, F(1, 381) = 2.139, p = 0.144. There were also no significant two-way interactions: data presentation approach did not interact with evidence strength (F(1, 381) = 2.220, p = 0.137) nor with motivation to elaborate (F(1, 381) = 0.291, p = 0.590; motivation to elaborate also did not interact with evidence strength (F(1, 381) = 0.055, p = 0.814).

Engagement scores significantly differed by level of motivation to elaborate (F(1, 381) = 78.975, p < 0.001), but not by data presentation approach (F(1, 381) = 2.277, p = 0.132) nor evidence strength (F(1, 381) = 0.027, p 0.869). Those with low levels of motivation to elaborate (M = 4.48, SD = 0.61, CI95%[4.36 to 4.60]) had significantly lower engagement scores than those with high levels of motivation to elaborate (M = 5.26, SD = 0.63, CI95%[5.14 to 5.38]; See Figure 27).

# Figure 27.





Attitude Change Analyses. In order to provide a nuanced and holistic examination of factors that impact attitude change, hierarchical multiple regression analyses were utilized to explore factors that impact global and evaluand-specific attitude change. In order to assess whether trends from Phase 1 and 2 were replicated with interactive data presentations (hypothesis 4.2), motivation to elaborate scores, dummy coded presentation approach and evidence strength level, and engagement scores were included, as well as two interaction terms

between (1) engagement scores and evidence strength and (2) presentation approach and evidence strength. Finally, to assess to whether the extent to which participants were surprised by the evaluation findings impacts global attitude change (hypothesis 4.3), participants' responses on a 7-point semantic differential scale to the question "To what extent, if at all, are the study findings surprising?" were also included in the regression analyses.

*Global Attitude Change*. Overall, participants tended to have slightly negative changes in their global attitude scores. Additionally, participants reported moderate levels of motivation to elaborate, moderate engagement with the presentation of evaluation findings, and were moderately surprised by the evaluation findings (see Table 62). Of the 433 participants included in the analyses, 219 were presented weak evidence while 214 were presented strong evidence. Similarly, 219 viewed a minimalist presentation of the findings while 214 received an interactive presentation of the findings.

Dependent Variable	N	Mean	SD	Min	Max	Skew	Kurtosis
Global Attitude Change	433	-0.15	0.87	-6.00	2.40	-1.59	6.49
Specific Attitude Change	433	-0.01	0.92	-6.00	4.40	-1.63	9.41
Motivation to Elaborate	433	5.02	1.51	1.00	7.00	-0.84	0.11
Engagement	433	4.90	0.92	1.75	7.00	-0.64	0.52
Surprise	433	4.50	1.70	1.00	7.00	-0.39	-0.90

Table 62.

Phase 3 participants' dependent and independent variable descriptive statistics.

There was independence of residuals (Durbin-Watson = 1.398). Examination of studentized residual plots and partial regression plots confirmed linear relationships between global attitude change, centered motivation to elaborate scores, centered engagement scores, and centered surprise scores. Motivation to elaborate scores, engagement scores, and surprise scores

were centered to reduce multicollinearity. Pearson's correlations indicated no multicollinearity (Pearson's |r|<0.7) and the minimum Tolerance statistic was 0.336.

A total of seven outliers were identified as having standardized residuals greater than three standard deviations. However, as no case had a leverage point greater than 0.2, these were retained in the analyses. Similarly, no cases had a Cook's Distance value above 1 (maximum Cook's Distance = 0.41). Examination of P-P Plot indicated approximately normally distributed residual errors.

The addition of interaction terms between engagement scores and evidence strength and between evidence strength and presentation approach led to a statistically significant increase in  $R^2$ , F(2, 425) = 5.773, p = 0.003. The full model of motivation to elaborate scores, surprise scores, engagement scores, presentation approach, evidence strength, and the interaction terms significantly predicted global attitude change,  $R^2 = 0.173$ , F(7, 425) = 12.728, p < 0.001, adjusted  $R^2 = 0.160$  (See Table 63).

#### Table 63.

Moderation effects of evidence strength on presentation approach predicting global attitude change.

Step	Variable	r	R <sup>2</sup> Change	В	SE <sub>B</sub>	β
1	Evidence Strength (Weak = 0; Strong=1)	.341***	0.139***	0.360**	0.108	0.207**
	Presentation Approach (Minimal=0; Interactive=1)	-0.129**	0.139***	-0.446**	0.109	-0.257**
	Motivation to Elaborate	-0.089*	0.139***	-0.079***	0.030	-0.136***
2	Engagement Score	0.017	0.011	0.189**	0.063	0.199**
	Surprise Score	-0.094*	0.011	-0.039	0.024	-0.076
3	Interaction Term 1: (Engagement Score x Evidence Strength)	-0.042	0.022**	-0.153	0.084	-0.110
	Interaction Term 2: (Presentation Approach x Evidence Strength)	0.189***	0.022**	0.454**	0.154	0.224**
(Consta	ant)			-0.221	0.076	

\*p<0.05; \*\*p < -.01; \*\*\*p <0.001; Cumulative  $R^2$ = 0.173; Adjusted  $R^2$ = 0.160; Motivation to Elaborate is 'centered' to a mean of 5.02, Engagement is 'centered' to a mean of 4.90, and Surprise is 'centered' to a mean of 4.40.

Examination of standardized beta coefficients indicated a significant interaction between presentation approach and evidence strength (t(425) = 2.944, p = 0.003) but not a significant interaction between engagement and evidence strength (t(425) = -1.821, p = 0.069. Figure 28 illustrates the interaction effect between presentation approach and evidence strength on predicted global attitude change, holding all other variables constant and utilizing the following regression equation:

# **Equation 6.**

Regression equation predicting global attitude change.

 $\hat{Y} = -0.221 + 0.360$  (Evidence Strength) - 0.446 (Presentation Approach)

- 0.079(Motivation to Elaborate) + 0.189 (Engagement Score)
- 0.039(Surprise Score)
- 0.153(Engagement Score x Evidence Strength)
- + 0.454(*Presentation Approach x Evidence Strength*)

### Figure 28.

Interaction effect between evidence strength and presentation approach on predicted global attitude change, holding motivation to elaborate, engagement scores, surprise scores constant.



**Note.** Figure holds constant motivation to elaborate scores, engagement scores, surprise scores (all at +1 standard deviation from the mean), and minimalist presentation approach for illustrative purposes. Figure is not a complete description of the data.

As seen in Table 63, the extent to which participants reported being surprised with the evaluation findings did not significantly predict global attitude change (t(425)= -1.639, p = 0.102). However, for participants who were shown interactive presentations of findings, a difference score was calculated between participants' guess of the evaluation findings and the actual evaluation finding. The average of these difference scores had only a weak correlation with surprise scores (Pearson's r = 0.137, p = 0.045; see Table 64). As a result, hierarchical multiple regression analyses predicting global attitude change were replicated for participants who were shown interactive presentations and replacing surprise scores with the average difference scores. Findings between the two approaches were consistent, and average difference scores did not significantly predict global attitude change (B = -0.004,  $\beta$  = -0.077, t(210) = -1.053, p = 0.293).

#### Table 64.

Descriptive statistics for difference between expected and actual evaluation findings.

Variable	N	Mean	SD	Min	Max	Skew	Kurtosis
Average difference score	215	11.02	18.29	-29.50	88.25	1.25	2.62

*Evaluand-specific Attitude Change*. A second hierarchical multiple regression analysis was conducted to explore impact of the same factors on evaluand-specific attitude change (See Table 62). There was independence of residuals (Durbin-Watson = 1.738). Examination of studentized residual plots and partial regression plots confirmed linear relationships between evaluand-specific attitude change, centered motivation to elaborate scores, centered engagement scores, and centered surprise scores.

There was homoscedasticity, as assessed by visual inspection of a plot of studentized residuals versus unstandardized predicted values. Motivation to elaborate scores, engagement scores, and surprise scores were centered to reduce multicollinearity. Pearson's correlations

indicated no multicollinearity (Pearson's |r|<0.7) and the minimum Tolerance statistic was 0.336.

A total of eight outliers were identified as having standardized residuals greater than three standard deviations. However, as no case had a leverage point greater than 0.2, these were retained in the analyses. Similarly, no cases had a Cook's Distance value above 1 (maximum Cook's Distance = 0.38). Examination of P-P Plot indicated approximately normally distributed residual errors.

The addition of interaction terms between engagement scores and evidence strength and between evidence strength and presentation approach led to a statistically significant increase in  $R^2$ , F(2, 425) = 3.129, p = 0.045. The full model of motivation to elaborate scores, surprise scores, engagement scores, presentation approach, evidence strength, and the interaction terms significantly predicted global attitude change,  $R^2 = 0.194$ , F(7, 425) = 14.630, p < 0.001, adjusted  $R^2 = 0.181$  (See Table 65).

Examination of standardized beta coefficients indicated a significant interaction between presentation approach and evidence strength (t(425) = 2.478, p = 0.014) but not a significant interaction between engagement and evidence strength (t(425) = -0.449, p = 0.653. Figure 29 illustrates the interaction effect between presentation approach and evidence strength on predicted global attitude change, holding all other variables constant and utilizing the following regression equation:

# **Equation 7.**

Regression equation predicting evaluand-specific attitude change.

 $\hat{Y} = -0.068 + 0.408$ (Evidence Strength) - 0.482(Presentation Approach)

- 0.130(Motivation to Elaborate) + 0.162(Engagement Score)

- 0.064(*Surprise Score*)

- 0.040(Engagement Score x Evidence Strength)

+ 0.402(Presentation Approach x Evidence Strength)

# Table 65.

Moderation effects of evidence strength on presentation approach predicting evaluand-specific attitude change.

Step	Variable	r	R <sup>2</sup> Change	В	SEB	β
1	Evidence Strength (Weak = 0; Strong=1)	0.336***	0.163***	0.408***	0.114	0.220***
	Presentation Approach (Trad. = 0; Interactive = 1)	-0.160***	0.163***	-0.482***	0.114	-0.261***
	Motivation to Elaborate	-0.163***	0.163***	-0.130***	0.032	-0.212***
2	Engagement Score	-0.017	0.020**	0.162*	0.067	0.160*
	Surprise Score	-0.151**	0.020**	-0.064*	0.025	-0.117*
3	Interaction Term (Engagement Score x Evidence Strength) Interaction Term	-0.022	0.012*	-0.040	0.089	-0.027
	(Presentation Approach x Evidence Strength)	0.153**	0.012*	0.402*	0.162	0.186*
(Consta	ant)			-0.068	0.080	

\*p < 0.05; \*\*p < -.01; \*\*\*p < 0.001; Cumulative  $R^2 = 0.194$ ; Adjusted  $R^2 = 0.181$ ; Motivation to Elaborate is 'centered' to a mean of 5.02, Engagement is 'centered' to a mean of 4.90, and Surprise is 'centered' to a mean of 4.40.

#### Figure 29.

Interaction effect between evidence strength and presentation approach on predicted evaluandspecific attitude change, holding motivation to elaborate, engagement scores, surprise scores constant.



**Note.** Figure holds constant motivation to elaborate scores, engagement scores, surprise scores (all at +1 standard deviation from the mean), and minimalist presentation approach for illustrative purposes. Figure is not a complete description of the data.

Additionally, to further examine the extent to which participants' expectations of the evaluation findings impact evaluand-specific attitude change, the analyses were replicated for participants who viewed the interactive presentation, replacing surprise scores with the average difference between participants' guess of the evaluation findings and the actual evaluation findings (see Table 64). Similarly to surprise scores, average difference scores significantly predicted evaluand-specific attitude change (B = -0.011,  $\beta = -0.192$ , t(214) = 0.010).

Attitude Strength Analyses. In order to examine factors that impact attitude strength, four hierarchical multiple regressions were conducted to separately examine effects on attitude resilience, attitude persistence (global and evaluand-specific), and behavioral intent (hypothesis 4.3; See Table 66). On average, Phase 3 participants showed greater evaluand-specific attitude

change after being shown a negative message about the program (M = -0.30, SD = 0.96) than after seeing the evaluation findings (M = -0.01, SD = 0.92; See Table 62).

# Table 66.

5.
5

Independent Variable	N	Mean	SD	Min	Max	Skew	Kurtosis
Attitude Resilience	171	-0.30	0.96	-2.40	2.60	0.69	0.19
Global Attitude Persistence	180	0.05	0.91	-3.80	3.00	0.01	3.63
Specific Attitude Persistence	180	0.01	0.83	-2.00	4.80	1.23	6.09
Behavioral Intent	433	4.97	1.53	1.00	7.00	-0.88	0.21

Attitude Resilience. Of the 176 participants recruited into the resilience subsample, 171

(97.2%) completed the five-item resilience questionnaire. To assess attitude resilience, a change score was calculated between post-intervention evaluand-specific attitudes and those after a negative message about the WISH program. A hierarchical multiple regression was conducted to determine factors predicting attitude resilience. Descriptive statistics of independent variables are summarized in Table 67 and Table 68.

# Table 67.

Phase 3 resilience subsample descriptive statistics of independent variables.

Independent Variable	N	Mean	SD	Min	Max	Skew	Kurtosis
Motivation to Elaborate	171	5.01	1.57	1.00	7.00	-0.75	-0.15
Engagement	171	4.84	0.92	1.90	6.95	-0.47	0.64
Surprise	171	4.33	1.81	1.00	7.00	-0.30	-1.08

Variable	Evidence	Strength	Presentatio	n Approach
Condition	Weak	Strong	Minimalist	Interactive
N	88	83	88	83

**Table 68.**Frequency of conditions in resilience subsample.

There was independence of residuals (Durbin-Watson = 2.142). Examination of studentized residual plots and partial regression plots confirmed linear relationships between attitude resilience, centered motivation to elaborate scores, centered engagement scores, and centered surprise scores. There was homoscedasticity, as assessed by visual inspection of a plot of studentized residuals versus unstandardized predicted values. Motivation to elaborate scores, engagement scores, and surprise scores were centered to reduce multicollinearity. Pearson's correlations revealed multicollinearity between surprise scores and the interaction term between engagement and evidence strength (r=0.708). As a result, surprise scores were removed from the analyses, and there was no multicollinearity detected between the remaining variables (Pearson's |r| < 0.7). The minimum Tolerance statistic was 0.347.

A total of three outliers were identified as having standardized residuals greater than three standard deviations. However, as no case had a leverage point greater than 0.2, these were retained in the analyses. Similarly, no cases had a Cook's Distance value above 1 (maximum Cook's Distance = 0.08). Examination of P-P Plot indicated approximately normally distributed residual errors.

The addition of interaction terms between engagement scores and evidence strength and between evidence strength and presentation approach did not lead to a statistically significant increase in  $R^2$ , F(2, 164) = 0.106, p = 0.899. None of the regression models produced significantly predicted attitude resilience (Final regression model F(6, 164) = 1.490, p=0.184;

See Table 69).

# Table 69.

*Regression model predicting evaluand-specific attitude resilience.* 

Step	Variable	r	$R^2$ Change	В	SE <sub>B</sub>	β
1	Evidence Strength (Weak = 0; Strong=1)	0.106	0.034	0.263	0.202	0.138
	Presentation Approach (Trad. = 0; Interactive = 1)	0.002	0.034	0.033	0.203	0.017
	Motivation to Elaborate	-0.152*	0.034	-0.135	0.053	-0.222
2	Engagement Score	0.037	0.017	0.157	0.121	0.151
3	Interaction Term (Engagement Score x Evidence Strength)	0.029	0.001	-0.007	0.158	-0.005
	Interaction Term (Presentation Approach x Evidence Strength)	0.037	0.001	-0.133	0.291	-0.059
(Constan	nt)			-0.422	0.142	

\*p < 0.05; \*\*p < -.01; \*\*\*p < 0.001; Cumulative  $R^2 = 0.041$ ; Adjusted  $R^2 = 0.006$ ; Motivation to Elaborate is 'centered' to a mean of 5.01, Engagement is 'centered' to a mean of 4.84.

Attitude Persistence. Two hierarchical multiple regressions were conducted on responses

of the 180 participants who responded to the follow-up survey after a week's delay. To

determine attitude persistence, a change score was calculated between post-intervention global

and evaluand-specific attitudes and those after a week delay. Descriptive statistics of persistence

subsamples' independent variables are provided in Table 70 and Table 71.

# Table 70.

Descriptive statistics of independent variables predicting global attitude persistence.

Independent Variable	N	Mean	SD	Min	Max	Skew	Kurtosis
Motivation to Elaborate	180	5.08	1.42	1.00	7.00	-1.03	0.75
Engagement Score	180	5.02	0.83	2.55	6.55	-0.59	0.08
Surprise Score	180	4.63	1.61	1.00	7.00	-0.35	-0.86

Variable	Evidence	Strength	Presentation	n Approach
Condition	Weak	Strong	Minimalist	Interactive
N	91	89	89	91

**Table 71.**Frequency of conditions in persistence subsample.

First, a hierarchical multiple regression was conducted to assess factors which predict global attitude persistence. There was independence of residuals (Durbin-Watson = 1.790). Examination of studentized residual plots and partial regression plots confirmed linear relationships between global attitude persistence, centered motivation to elaborate scores, centered engagement scores, and centered surprise scores. There was homoscedasticity, as assessed by visual inspection of a plot of studentized residuals versus unstandardized predicted values. Motivation to elaborate scores, engagement scores, and surprise scores were centered to reduce multicollinearity. Pearson's correlations indicated no multicollinearity (Pearson's |r|<0.7) and the minimum Tolerance statistic was 0.328.

A total of three outliers were identified as having standardized residuals greater than three standard deviations. However, as no case had a leverage point greater than 0.2, these were retained in the analyses. Similarly, no cases had a Cook's Distance value above 1 (maximum Cook's Distance = 0.15). Examination of P-P Plot indicated approximately normally distributed residual errors.

The addition of interaction terms between engagement scores and evidence strength and between evidence strength and presentation approach did not lead to a statistically significant increase in  $R^2$ , F(2, 172) = 0.252, p = 0.777. The full model of motivation to elaborate scores, surprise scores, engagement scores, presentation approach, evidence strength, and the interaction

155

terms significantly predicted global attitude persistence,  $R^2 = 0.111$ , F(7, 172) = 3.066, p = 0.005, adjusted  $R^2 = 0.075$  (See Table 72).

# Table 72.

Final regression model predicting global attitude persistence.

Step	Variable	r	R <sup>2</sup> Change	В	SEB	β
1	Evidence Strength (Weak = 0; Strong=1)	-0.282***	0.107***	-0.389*	0.190	-0.214
	Presentation Approach (Trad. = 0; Interactive = 1)	0.160*	0.107***	0.358	0.188	0.197
	Motivation to Elaborate	0.105	0.107***	0.075	0.057	0.117
2	Engagement Score	0.039	0.002	-0.063	0.121	-0.058
	Surprise Score	-0.003	0.002	-0.012	0.044	-0.021
3	Interaction Term (Engagement Score x Evidence Strength) Interaction Term	0.048	0.003	0.044	0.161	0.027
	(Presentation Approach x Evidence Strength)	-0.093	0.003	-0.182	0.270	-0.085
(Constant)				0.102	0.137	

\*p<0.05; \*\*p < -.01; \*\*\*p <0.001; Cumulative  $R^2$ = 0.111; Adjusted  $R^2$ = 0.075; Motivation to Elaborate is 'centered' to a mean of 5.08, Engagement is 'centered' to a mean of 5.02, and Surprise is 'centered' to a mean of 4.63.

As summarized in Table 72, examination of standardized beta coefficients revealed

evidence strength significantly predicted global attitude persistence, (t(172) = -2.046, p = 0.042).

Figure 30 illustrates the effect of evidence strength on predicted global attitude persistence

scores, holding all other variables constant and utilizing the following regression equation:

## **Equation 8.**

Regression equation predicting global attitude persistence.

 $\hat{Y} = 0.102 - 0.389$ (Evidence Strength) + 0.358(Presentation Approach)

+ 0.075(*Motivation to Elaborate*) – 0.063(*Engagement Score*)

- 0.012(Surprise Score)
- 0.044(Engagement Score x Evidence Strength)
- 0.182(Presentation Approach x Evidence Strength)

#### Figure 30.

Effect of evidence strength on predicted global attitude persistence scores, holding motivation to elaborate scores, engagement scores, surprise scores, interactive presentation, and strong evidence constant.



Predicted Global Attitude Persistence

**Note.** Scores closer to 0 indicate greater global attitude persistence. Figure holds constant motivation to elaborate scores, engagement scores, surprise scores (all at +1 standard deviation from the mean), and interactive presentation approach for illustrative purposes. Figure is not a complete description of the data.

The second hierarchical multiple regression assessed the impact of factors on evaluandspecific attitude persistence. There was independence of residuals (Durbin-Watson = 1.977). Examination of studentized residual plots and partial regression plots confirmed linear relationships between evaluand-specific attitude persistence, centered motivation to elaborate scores, centered engagement scores, and centered surprise scores. There was homoscedasticity, as assessed by visual inspection of a plot of studentized residuals versus unstandardized predicted values. Motivation to elaborate scores, engagement scores, and surprise scores were centered to reduce multicollinearity. Pearson's correlations indicated no multicollinearity (Pearson's |r|<0.7) and the minimum Tolerance statistic was 0.328. A total of one outlier was identified as having standardized residuals greater than three standard deviations. However, as no case had a leverage point greater than 0.2, these were retained in the analyses. Similarly, no cases had a Cook's Distance value above 1 (maximum Cook's Distance = 0.13). Examination of P-P Plot indicated approximately normally distributed residual errors.

The addition of interaction terms between engagement scores and evidence strength and between evidence strength and presentation approach did not lead to a statistically significant increase in  $R^2$ , F(2, 172) = 0.002, p = 0.998. The full model of motivation to elaborate, surprise score, engagement score, presentation approach, evidence strength, and the interaction terms significantly predicted evaluand-specific attitude persistence,  $R^2 = 0.079$ , F(7, 172) = 2.096, p =0.046, adjusted  $R^2 = 0.041$  (See Table 73).

Examination of standardized beta coefficients revealed evidence strength significantly predicted evaluand-specific attitude persistence, (t(172) = -2.101, p = 0.037). Figure 31 illustrates the effect of evidence strength on predicted evaluand-specific attitude persistence scores, holding all other variables constant and utilizing the following regression equation:

# **Equation 9.**

Regression equation predicting evaluand-specific attitude persistence.

 $\widehat{Y} = 0.082 - 0.371$  (Evidence Strength) + 0.229 (Presentation Approach)

- + 0.031(Motivation to Elaborate) 0.092(Engagement Score)
- 0.005(Surprise Score)
- 0.007(Engagement Score x Evidence Strength)
- 0.011(Presentation Approach x Evidence Strength)

# Table 73.

Step	Variable	r	$R^2$ Change	В	SE <sub>B</sub>	β
1	Evidence Strength (Weak = 0; Strong=1)	-0.237**	0.072**	-0.371*	0.177	-0.224*
	Presentation Approach (Trad. = 0; Interactive = 1)	0.141*	0.072**	0.229	0.175	0.138
	Motivation to Elaborate	0.015	0.072**	0.031	0.053	0.053
2	Engagement Score	-0.051	0.007	-0.092	0.112	-0.092
	Surprise Score	-0.016	0.007	-0.005	0.041	-0.010
3	Interaction Term (Engagement Score x Evidence Strength)	-0.033	0.000	-0.007	0.149	-0.005
	Interaction Term (Presentation Approach x Evidence Strength)	-0.056	0.000	-0.011	0.250	-0.006
(Constant)				0.082	0.128	

Final regression model predicting evaluand-specific attitude persistence.

\*p<0.05; \*\*p < -.01; \*\*\*p <0.001; Cumulative  $R^2$ = 0.079; Adjusted  $R^2$ = 0.041; Motivation to Elaborate is 'centered' to a mean of 5.08, Engagement is 'centered' to a mean of 5.02, and Surprise is 'centered' to a mean of 4.63.

# Figure 31.

Effect of evidence strength on predicted evaluand-specific attitude persistence, holding motivation to elaborate scores, engagement scores, surprise scores, and interactive presentation constant.



Predicted WISH-specific Attitude Persistence

**Note.** Scores closer to 0 indicate greater evaluand-specific attitude persistence. Figure holds constant motivation to elaborate, engagement scores, surprise scores (both at +1 standard deviation from the mean), and interactive presentation approach for illustrative purposes. Figure is not a complete description of the data.

*Behavioral Intent.* To assess behavioral intent, an average score of four behavioral intent items was calculated (see Table 66). A hierarchical multiple regression was conducted to assess what factors impact behavioral intent.

There was independence of residuals (Durbin-Watson = 1.798). Examination of studentized residual plots and partial regression plots confirmed linear relationships between behavioral intent scores, centered motivation to elaborate scores, centered engagement scores, and centered surprise scores. There was homoscedasticity, as assessed by visual inspection of a plot of studentized residuals versus unstandardized predicted values. Motivation to elaborate scores, engagement scores, and surprise scores were centered to reduce multicollinearity. Pearson's correlations indicated no multicollinearity (Pearson's |r|<0.7) and the minimum Tolerance statistic was 0.336.

A total of seven outliers were identified as having standardized residuals greater than three standard deviations. However, as no case had a leverage point greater than 0.2, these were retained in the analyses. Similarly, no cases had a Cook's Distance value above 1 (maximum Cook's Distance = 0.08). Examination of P-P Plot indicated approximately normally distributed residual errors.

The addition of interaction terms between engagement scores and evidence strength and between evidence strength and presentation approach did not lead to a statistically significant increase in  $R^2$ , F(2, 425) = 0.894, p = 0.410. The full model of motivation to elaborate scores, surprise scores, engagement scores, presentation approach, evidence strength, and the interaction terms significantly predicted behavioral intent,  $R^2 = 0.467$ , F(7, 425) = 53.626, p < 0.001, adjusted  $R^2 = 0.460$  (See Table 74).

160

Examination of standardized beta coefficients revealed the following three factors

significantly predicted behavioral intent in order of effect size: motivation to elaborate (t(425) =

12.904, p < 0.001), engagement scores (t(425) = 4.034, p < 0.001), and evidence strength (t(425))

= 2.422, p = 0.016). Figure 32 through Figure 34 illustrate the impact of these three factors on

predicted behavioral intent scores, holding all other variables constant and utilizing the following

regression equation:

# **Equation 10.**

Regression equation predicting behavioral intent.

 $\widehat{Y} = 4.854 + 0.370$  (Evidence Strength) - 0.279 (Presentation Approach)

- + 0.550(Motivation to Elaborate) + 0.361(Engagement Score)
- 0.048(*Surprise Score*)
- 0.033(Engagement Score x Evidence Strength)
- + 0.287(*Presentation Approach x Evidence Strength*)

# Table 74.

Final regression model predicting behavioral intent.

Step	Variable	r	R <sup>2</sup> Change	В	SE <sub>B</sub>	β
1	Evidence Strength (Weak = 0; Strong=1)	0.153**	0.438***	0.370*	0.153	0.121*
	Presentation Approach (Trad. = 0; Interactive = 1)	-0.058	0.438***	-0.279	0.154	-0.091
	Motivation to Elaborate	0.639***	0.438***	0.550***	0.043	0.542***
2	Engagement Score	0.473***	0.029***	0.361***	0.089	0.216***
	Surprise Score	0.113**	0.029***	-0.048	0.034	-0.053
3	Interaction Term (Engagement Score x Evidence Strength) Interaction Term	0.326***	0.002	-0.033	0.119	-0.013
	(Presentation Approach x Evidence Strength)	0.088*	0.002	0.287	0.217	0.080
(Constant)			4.854	0.108		

\*p<0.05; \*\*p < -.01; \*\*\*p <0.001; Cumulative  $R^2$ = 0.467; Adjusted  $R^2$ = 0.460; Motivation to Elaborate is 'centered' to a mean of 5.02, Engagement is 'centered' to a mean of 4.90, and Surprise is 'centered' to a mean of 4.40.
# Figure 32.

*Effect of motivation to elaborate on predicted behavioral intent scores, holding engagement scores, surprise scores, interactive presentation, and strong evidence constant.* 



**Note.** Figure holds constant engagement scores, surprise scores (both at +1 standard deviation from the mean), interactive presentation approach, and strong evidence for illustrative purposes. Figure is not a complete description of the data.

# Figure 33.

Effect of engagement scores on predicted behavioral intent scores, holding of motivation to elaborate, surprise scores, interactive presentation, and strong evidence constant.



**Note.** Figure holds constant motivation to elaborate scores, surprise scores (both at +1 standard deviation from the mean), interactive presentation approach, and strong evidence for illustrative purposes. Figure is not a complete description of the data.

### Figure 34.

Effect of evidence strength on predicted behavioral intent scores, holding motivation to elaborate scores, engagement scores, surprise scores, and interactive presentation constant.



**Note.** Figure holds constant motivation to elaborate scores, engagement scores, surprise scores (all at +1 standard deviation from the mean), and interactive presentation approach for illustrative purposes. Figure is not a complete description of the data.

Donation Behavior Analysis. Of the 433 participants, 113 (26.1%) decided to donate some or

all their HIT earnings to the WISH program. For those that chose not to donate, their donation

amount was marked as \$0.00. Average donation amounts for Phase 3 participants can be seen in

Table 75.

## Table 75.

Phase 3 participants' donation amount descriptive statistics.

Independent Variable	N	Mean	SD	Min	Max	Skew	Kurtosis
Donation Amount	433	0.20	0.50	0.00	2.40	3.23	10.31

Hierarchical multiple regression analyses were conducted to assess factors that predict donation amounts. In order to assess the extent to which attitudes predicted donation behaviors, post-intervention evaluand-specific attitudes were included in the model. There was independence of residuals (Durbin-Watson = 1.996). Examination of studentized residual plots and partial regression plots confirmed linear relationships between donation amount, centered motivation to elaborate scores, centered engagement scores, centered surprise scores, and center post-intervention evaluand-specific attitudes. There was homoscedasticity, as assessed by visual inspection of a plot of studentized residuals versus unstandardized predicted values. Motivation to elaborate scores, engagement scores, surprise scores and post-intervention evaluand-specific attitudes were centered to reduce multicollinearity. Pearson's correlations indicated no multicollinearity (Pearson's |r|<0.7) and the minimum Tolerance statistic was 0.334.

A total of 16 outliers were identified as having standardized residuals greater than three standard deviations. However, as no case had a leverage point greater than 0.2, these were retained in the analyses. Similarly, no cases had a Cook's Distance value above 1 (maximum Cook's Distance = 0.06). Examination of P-P Plot indicated approximately normally distributed residual errors.

The addition of interaction terms between engagement scores and evidence strength and between evidence strength and presentation approach did not lead to a statistically significant increase in  $R^2$ , F(2, 424) = 1.706, p = 0.183. The full model of motivation to elaborate scores, surprise score, engagement score, post-intervention evaluand-specific attitudes, presentation approach, evidence strength, and the interaction terms significantly predicted donation amount,  $R^2 = 0.100$ , F(8, 424) = 5.901, p < 0.001, adjusted  $R^2 = 0.083$  (See Table 76).

### Table 76.

Step	Variable	r	R <sup>2</sup> Change	В	SEB	β
1	Evidence Strength (Weak = 0; Strong=1)	0.050	0.043***	0.036	0.066	0.036
	Presentation Approach (Trad. = 0; Interactive = 1)	-0.016	0.043***	-0.116	0.066	-0.115
	Motivation to Elaborate	0.199***	0.043***	0.049	0.019	0.146*
2	Engagement Score	0.225***	0.050***	0.126**	0.038	0.230**
	Surprise Score	0.129**	0.050***	0.010	0.015	0.035
	Evaluand-specific Post- attitudes	-0.070***	0.050***	-0.078***	0.022	-0.186***
3	Interaction Term (Engagement Score x Evidence Strength) Interaction Term	0.124**	0.007	-0.055	0.051	-0.068
	(Presentation Approach x Evidence Strength)	0.054	0.007	0.144	0.093	0.123
(Const	ant)			0.204	0.046	

Final regression model approach predicting donation amounts.

\*p<0.05; \*\*p < -.01; \*\*\*p < 0.001; Cumulative  $R^2 = 0.073$ ; Adjusted  $R^2 = 0.057$ ; Motivation to Elaborate is 'centered' to a mean of 5.02, Engagement is 'centered' to a mean of 4.90, Surprise is 'centered' to a mean of 4.40, and Post-intervention evaluand-specific attitudes are 'centered' to a mean of 5.83.

Examination of standardized beta coefficients indicated a significant main effect of

engagement scores (t(425) = 22.837, p = 0.005), post-intervention evaluand-specific attitudes

(t(425) = -3.602, p < 0.001), and motivation to elaborate (t(425) = 2.594, p = 0.010). Figure 35

through Figure 37 illustrate these effects on predicted donation behavior, holding all other

variables constant and utilizing the following regression equation:

### **Equation 11.**

Regression equation predicting donation amount.

 $\hat{Y} = 0.204 + 0.036$  (Evidence Strength) - 0.116 (Presentation Approach)

- + 0.049(*Motivation to Elaborate*) + 0.126(*Engagement Score*)
- + 0.010(Surprise Score)
- 0.078 (WISH specific postintervention attitudes)
- 0.055(Engagement Score x Evidence Strength)
- + 0.144(Presentation Approach x Evidence Strength)

# Figure 35.

*Effect of engagement on predicted donation amounts, holding motivation to elaborate, surprise scores, post-intervention attitudes, interactive visualization with strong evidence constant.* 



**Note.** Figure holds constant motivation to elaborate scores, surprise scores, and post-intervention evaluand-specific attitudes (all at +1 standard deviation from the mean), and interactive presentation with strong evidence for illustrative purposes. Figure is not a complete description of the data.

## Figure 36.

Effect of post-intervention evaluand-specific attitudes on predicted donation amounts, holding motivation to elaborate scores, surprise scores, engagement scores, interactive visualization with strong evidence constant.



**Note.** Figure holds constant motivation to elaborate scores, surprise scores, and engagement scores (all at +1 standard deviation from the mean), and interactive presentation with strong evidence for illustrative purposes. Figure is not a complete description of the data.

#### Figure 37.

Effect of motivation to elaborate on predicted donation amounts, holding post-intervention evaluand-specific attitudes, surprise scores, engagement scores, interactive visualization with strong evidence constant.



**Note.** Figure holds constant post-intervention evaluand-specific attitudes, surprise scores, and engagement scores (all at +1 standard deviation from the mean), and interactive presentation with strong evidence for illustrative purposes. Figure is not a complete description of the data.

#### DISCUSSION

The final phase of this research built upon lessons learned from the first two studies to continue to explore whether evaluation products could be designed to promote evaluation influence pathways. Specifically, the third phase of research explored the impact of interactive presentations of evaluation findings and the role of pre-existing expectations on attitudes and donation behaviors. Study findings indicate that interactive presentations of findings promote elaboration and impact attitude change, but not attitude strength and donation behaviors. Additionally, pre-existing expectations impact evaluand-specific attitude change.

Proponents of interactivity pose that the approach turns the audience from "a passive consumer into an active participant" and that it is a tool for increased transparency (Weissgerber, et al., 2016; Weissgerber et al., 2017, p. 20592). Study findings did not find evidence to support the first claim, which informed hypothesis 4.1, as no differences were found in participants'

reported engagement for interactive compared with static minimalist presentations of findings. As in prior phases of this research, the characteristics of the audience – specifically, motivation to elaborate – drove the extent to which participants found the presentations of the evaluation findings engaging, which reinforces the importance of considering audience when designing evaluation products.

The second claim – that interactivity is a tool for increased transparency – is more complex. While transparency was not directly measured, study findings suggest that interactivity facilitated elaboration on the evaluation findings and resulted in the differentiation of attitude change based on the strength of the evidence, consistent with hypothesis 4.2. This finding aligns with research on the Elaboration Likelihood Model (ELM), which indicates that one way to increase the likelihood of elaboration is to increase the audiences' involvement (Johnson, & Eagly, 1989; Petty & Cacioppo, 1986; Wang, Wang, & Farn, 2009). Interestingly, participants had similar attitude change when evaluation findings indicated strong effects of the evaluand; however, those that were shown interactive presentations of the evaluation findings had significantly more negative reactions to weak findings than those who were shown static minimalist data visualizations. Whether this trend is reflective of greater transparency is unclear - on one hand, interactivity seems to highlight the strength of the evaluation findings to the audience. However, it is also possible that the approach may bias participants, causing more extreme reactions than they would have had otherwise. Further research is needed to clarify the influence of interactive presentations of evaluation findings on participants' judgements of social programs.

Mark and Henry (2004) identify elaboration and attitude valence as both evaluation influence mechanisms and outcomes that can be triggered by evaluation products. However, as

noted by Herbert (2014), there is a lack of operationalization that would guide how evaluation products must be designed to produce these effects. This research contributes one small step towards closing this gap by demonstrating that interactive presentations of findings outperform static minimalist data visualizations in facilitating elaboration and attitude change.

However, from a practice perspective, the extent to which interactive presentations outperformed minimalist visualizations was relatively small. Interactivity provided an enhancement to the evaluation findings' influence, but other factors such as evidence strength and motivation to elaborate also pay a role. For evaluand-specific attitudes, prior expectations were found to further influence attitude change, partially supporting hypothesis 4.3 and in alignment with previous ELM research (Zanna & Cooper, 1974; Petty, Fabrigar, & Wegener, 2003). Additionally, the type of interactive presentation of findings explored in this study is not feasible in all evaluations – often, evaluation contracts dictate a static report as the evaluation product to be generated. Thus, evaluation practitioners may consider interactive presentations of findings as an opportunity to enhance their evaluations rather than a critical component of all evaluations. For example, shifting evaluation reports from static to web-based formats may allow evaluators to incorporate greater interactivity while meeting contract terms.

The impacts of interactivity were also limited to immediate attitude change and did not predict attitude strength and donation behaviors, providing evidence against hypotheses 4.4 and 4.5. These finding are not aligned with previous research on the Elaboration Likelihood Model (ELM) as elaboration is associated with greater attitude strength and behavior predictivity (e.g., Crano, 1995; Petty & Cacioppo, 1986). Instead, attitude persistence was predicted by the strength of the evaluation findings, regardless of how they were presented. For both global and evaluand-specific attitude persistence, strong evidence was associated with slightly less positive

attitudes after a week's delay as compared to immediately after seeing the evaluation findings while weak evidence tended to have significantly more positive attitudes after a week's delay than immediately after seeing the evaluation findings. Together, these findings suggest that participants' attitudes tended to regress over time, more closely reflecting their initial attitudes prior to reading the evaluation findings; however, this trend was amplified for weak evaluation findings.

These findings have implications for evaluation practice. Evaluations are likely to be less impactful if there are delays between when evaluation findings are presented and the decisions they are meant to inform. It may be beneficial to either time the presentation of findings to align with key decision points or to provide stakeholders with reminders. Indeed, evaluation finding reminders may be particularly helpful, as ELM research has found that repetition increases the ability to elaborate on findings (e.g., Haugtvedt, Schumann, Schneier, & Warren, 1994; Lein, 2001). Practices that reinforce participants' knowledge of evaluation findings over time may be particularly important when findings are weak or negative. Participants tended to react negatively to weak findings but returned to more positive assessments of the evaluand over time, suggesting a reduction in evaluation influence.

Replicating Phase 2 findings, none of the explored factors predicted attitude resilience. Like attitude persistence, this finding is surprising given previous research on ELM which indicates that elaboration is associated with greater attitude resilience (Petty & Cacioppo, 1986). Further research is needed to determine factors that impact attitude resilience within an evaluation context.

Intention to support the evaluand was primarily driven by the audience's characteristics – specifically, motivation to elaborate. In previous research on the ELM, central route processing

through elaboration was associated with greater behavioral intent (Crano, 1995). Current study findings suggest that while the presentation approach influenced elaboration sufficiently to impact attitude change, initial level of motivation to elaborate largely determined whether an individual would express intention to support the evaluand. However, engagement with the data presentation and the strength of the evaluation findings also impacted intent to support the program, though to a lesser extent. These findings align with the Theory of Planned Behavior, which highlights that attitudes are only one of several components that inform behavioral intention (Ajzen, 1991). As a result, behavioral intention may be more difficult to influence than attitudes.

Actual donation behaviors, however, were primarily driven by the extent to which participants engaged with the data presentation, regardless of the evidence strength. Based on ELM research, this suggests a more peripheral route, where participants' behaviors were influenced more by peripheral factors rather than the arguments – the evaluation findings – presented in favor of the evaluand (Petty & Cacioppo, 1986). However, there may also be a third factor that accounts for both the donation behavior and engagement trends. For example, the description of this research indicated that it was about a non-profit. It may be that those who volunteered for the study value non-profits, and that this characteristic drove both their engagement with the presentation and donation behaviors. Further research is needed to clarify the role of engagement with the presentations of evaluation findings and subsequent behaviors in support of the evaluand.

Donation amounts were also driven by post-intervention evaluand-specific attitudes, though to a lesser degree than presentation engagement. As in Phase 2, this finding aligns with evaluation influence pathways (Mark & Henry, 2004). In this case, the pathway began at the

evaluation product (data presentation) to a general influence mechanism of elaboration to a cognitive and affective mechanism of attitude valence and finally to an individual behavior change, tracing the full hypothesized pathway illustrated in Figure 4.

Overall, findings from all three phases of research indicate that evaluation influence is a complex process which benefits from careful consideration of the strength of evaluation findings, evaluation product design, and the characteristics and experiences of the audience.

## **CHAPTER 5 DISCUSSION**

The purpose of this research is to explore whether evaluation products whose design is informed by the Elaboration Likelihood Model (ELM) can trigger the creation of evaluation influence pathways (Mark & Henry, 2004; Petty & Cacioppo, 1986). The research examined three design approaches to presenting evaluation findings – minimalist data visualization, embellished data visualization, and interactive data presentations – and assessed the impacts of these products on audiences' experiences, attitudes, and donation behaviors. The following section summarizes key findings, research limitations, and implications for theory and practice.

#### **Interpretation of Results**

Findings from across the three phases of this research provide evidence that the design of evaluation products can trigger evaluation influence by facilitating elaboration on evaluation findings. However, not all design conventions produced these effects: only interactive data presentations impacted elaboration and consequent attitude change while minimalist and embellished data visualizations did not differ in their ability to produce evaluation influence. These findings are consistent with the principles of the Elaboration Likelihood Model (ELM), which indicates that greater personal involvement facilitates motivation to elaborate (Johnson, & Eagly, 1989; Petty & Cacioppo, 1986; Wang, Wang, & Farn, 2009). Informed by this principle, interactive data presentations provide the greatest level of audience involvement and were found to facilitate forming attitudes based on the strength of the evaluation findings.

However, the impact of the design of the evaluation products was limited to attitude change and did not affect attitude strength and donation behaviors. Indeed, factors that impact attitude persistence and resilience within an evaluation context remain largely unknown. The final component of attitude strength, behavioral intention, was primarily driven by the audience's

characteristics – in particular, their initial motivation to elaborate based on the relevance of the evaluand to their life and values. Similarly, donation behaviors were driven by motivation to elaborate as well as evaluand-specific attitudes formed after viewing the evaluation findings – regardless of the strength of the findings or the presentation approach utilized. Together, these findings provide evidence for evaluation influence pathways aligned with Mark and Henry's (2004) Evaluation Influence framework – that evaluation products facilitate elaboration which impacts the formation of attitudes which in turn influence donation behaviors.

In addition, the research findings highlight that audience characteristics, such as motivation to elaborate, play a significant role in almost all experiences and outcomes explored in this research. In terms of experience of the evaluation product, motivation to elaborate drove both engagement as well as satisfaction with the product and accuracy of interpretation. Individuals who were highly motivated to elaborate tended to be more engaged and satisfied with the presentations of the evaluation findings – regardless of presentation approach. Interestingly, however, high levels of motivation to elaborate also carried some drawbacks – these individuals tended to have lower interpretation accuracy and poor recall of the evaluation study. Additionally, individuals who were highly motivated to elaborate tended to spend more time examining findings that may have been surprising – highly motivated to elaborate individuals spent more time examining weak findings while those low in motivation to elaborate paused more when presented strong findings.

Motivation to elaborate continued to play a significant role in attitude change even when accounting for elaboration processes facilitated by the design of the evaluation product.

Interestingly, it was those who were low in motivation to elaborate that tended to have greater shifts in their attitudes than those who were highly motivated to elaborate. However, these trends did not transfer to behavioral intention and donation amounts. Those who were more motivated to elaborate tended to have more intention to support the evaluand, which translated to donating larger amounts. Indeed, motivation to elaborate had a larger influence on the amount donated than post-intervention, evaluand-specific attitudes. This suggests that motivation to elaborate may reflect non-attitudinal factors that influence behavior, such as social norms (Ajzen, 1991).

#### Limitations

While each phase of this research noted unique limitations, it is important to highlight the overarching limitations of this research study. In particular, while the experimental design isolated the impact of the factors explored in this research, it also limited external validity of the findings. For example, by study design, all participants read the evaluation findings, which helped facilitate assessment of how different groups process evaluation findings. However, outside of the study context, there may be differences in who voluntarily reads evaluation reports. In particular, low motivation to elaborate individuals were found to have greater attitude change than those who were highly motivated to elaborate on evaluation findings. However, had these individuals not been prompted to read the evaluation findings as part of the study task, they may have not read the report in the first place.

Similarly, this research also focused on presenting limited evaluation findings where a realworld evaluation would likely explore additional aspects of the evaluand, such as for whom it was most effective and program components that best supported outcomes (Gargani & Donaldson, 2011). Results may also have been different if a greater number and/or variety of evaluation findings were presented. In addition to limiting external validity, the presentation of

limited findings may have also reduced the strength of the evidence strength manipulation. Previous research on the Elaboration Likelihood Model piloted a wide variety of arguments to determine "weak" and "strong" arguments (Petty & Cacioppo, 1986). In the current research, the strength of the argument was determined by piloting different effect sizes of evaluation results but did not explore different outcomes that the evaluation study could have included. It is possible, therefore, that different outcomes than those explored could have resulted in stronger manipulations for the evidence strength factor explored.

The audience included in the study is limited to members of the public, which represents only one of several evaluation stakeholder groups (Weiss, 1998; Bryson, Patton, & Bowman, 2011). While participants who were highly motivated to elaborate included those who could potentially benefit from the evaluand if it had been included in their community, it did not include the typical primary evaluation stakeholders – program managers and decision-makers (Weiss, 1998). Research on the Elaboration Likelihood Model (ELM) defines individual involvement as the personal commitment a viewer may have to the issue at hand, which increases the likelihood of elaboration (Wang, Wang, & Farn, 2009). By this definition, program managers and decision-makers are likely to have high involvement given their commitment to the evaluand, which should result in high levels of motivation to elaborate. Given this, it is possible that evaluation products designed to further enhance involvement – such as interactive data presentations – may not be as impactful for these individuals.

This research is also limited to one possible evaluation product – the presentation of evaluation findings. However, as noted by Mark and Henry (2004), there are other evaluation outputs, such as reports, presentations, and briefings that were not included in this study. Additionally, the current study only explores one level of analysis of Mark and Henry's (2004)

Evaluation Influence framework: that of the individual. Further research is needed to explore how evaluation product design impacts pathways formed at the interpersonal and collective levels as well as motivational mechanisms of influence. This research also only explores donation behaviors, and does not examine other relevant behaviors that could be impacted by an evaluation, such as implementation practices by program practitioners or funding decisions by program funders. Further research is needed to understand how evaluation products impact key behaviors that reflect and promote evaluation influence.

Finally, the embellished visualizations utilized in this research may have biased study participants through the include of components intended to communicate a value message, such as smiling and frowning faces. Rather than responding to the data itself, it is possible that those low in motivation to elaborate were biased by these components rather than the evaluation findings themselves. Further research is needed with neutral embellished data visualization to determine the potential for bias with this data visualization approach.

#### **Theoretical Implications**

The findings of this research have theoretical implications not only for evaluation influence theory, but also for data visualization research and persuasion research focused on the Elaboration Likelihood Model (ELM). This section address each of these areas.

**Evaluation Influence.** Herbert (2014) found that the lack of operationalization of evaluation influence mechanisms in Mark and Henry's (2004) Evaluation Influence framework limited its utility in both research and practice. The current research takes the first steps at closing these gaps in three ways: (1) by uncovering evaluation product design features that can trigger evaluation influence mechanisms, (2) by empirically measuring three mechanisms of

evaluation influence at the individual level, and (3) by empirically linking these mechanisms in evaluation influence pathways.

Previous efforts to operationalize evaluation influence mechanisms did so by defining these mechanisms as outcomes (Oliver, 2008). However, there remained a gap as to how these outcomes are triggered – in short, what features of evaluation products facilitated evaluation influence. Given that presentations of findings are a common evaluation output, the current research explored three design approaches: minimalist data visualizations, embellished data visualizations, and interactive data presentations. Findings indicated that the design of the product did impact subsequent evaluation influence, with interactive presentations outperforming other approaches. This adds to the evaluation influence literature by providing empirical support of the relationship between evaluation products and evaluation influence as well as providing insights into considerations for design decisions.

The current research also empirically measures two mechanisms of evaluation influence: (1) the general influence mechanism of elaboration and (2) the attitudinal mechanism of attitude valence. This contribution may help future researchers and theorists operationalize these mechanisms within an evaluation context. Similarly, in measuring the links between evaluation products, elaboration, attitudes, and donation behaviors, the current research provides an empirical support for evaluation influence pathways. In light of limited evaluation use found by Fleischer and Christie (2009), these findings suggest that evaluation influence may be more farreaching, but that evaluation pathways may make it more difficult for evaluators to track this influence. These findings also emphasize the need to broaden evaluation outcomes beyond instrumental use to capture an evaluation's impact.

Additionally, the current research demonstrates that research from other fields such as persuasion psychology can help provide insights and guidance into what product design features may be effective at producing evaluation influence (Fleming, 2011). In this research, the Elaboration Likelihood Model augmented the Mark and Henry (2004) Evaluation Influence framework to trigger a particular influence pathway (Petty & Cacioppo, 1986). The combining of the ELM and the Evaluation Influence framework also provides theoretical grounding for how certain evaluation practices may support greater evaluation influence. For example, participatory approaches to evaluation increase stakeholders' involvement in evaluation (Weiss, 1998). Based on the ELM, participatory approaches may increase stakeholders' motivation to elaborate, facilitating greater evaluation influence. Similarly, Fleming (2011) noted that aspects of Utilization-focused evaluation (UFE) that increase the "personal factor", such as actively educating and involving intended users in interpreting findings and selecting intended users who value and care about evaluation findings, are likely to increase ability and motivation to elaborate on an evaluation. While further study is needed to test whether ELM principles underly UFE practices, the current study provides preliminary evidence of these underlying mechanisms in evaluation, taking the initial steps to move evaluation theory from a prescriptive approach to a descriptive one.

Additionally, this approach can be replicated with other theories tied to the general influence, cognitive and attitudinal, motivational, and behavioral mechanisms identified in Mark and Henry's (2004) Evaluation Influence framework to translate mechanisms into concrete guidance for the design of evaluation products. However, as discussed further below, not all principles of the ELM produced expected results. These findings suggest that further research is needed to confirm that principles from other fields apply to the evaluation context.

Finally, the research findings emphasize the importance of context for evaluation influence. In this instance, the audience's motivation to elaborate played a significant role in both experiences of evaluation and their associated outcomes. These findings suggest that the evaluation influence mechanisms may interact with external factors, impacting the extent to which the evaluation findings can produce influence. Expanding the framework to identify and include these external factors may facilitate tailoring of evaluation influence to specific contexts.

**Data Visualization.** Recent debates amongst researchers of data visualization have called into question most effective data visualization principles. The current research adds to this conversation insights on how three different conventions for presenting data perform on various metrics assessing audiences' experiences and outcomes. The findings of this research pose additional criteria for assessing the effectiveness of data presentations.

Previous research on data visualization used the criteria of speed and accuracy of interpretation to judge the effectiveness of a visualization. Proponents of minimalist visualization approaches argued that high data-ink ratios and elimination of "chart junk" facilitated both speed and accuracy of interpretation (Tufte, 1983). However, more recently, proponents of embellished and interactive presentations argue that these features offer additional benefits, such as increased engagement, memorability, and transparency (Bateman et al., 2010; Byrne, Angus, & Wiles, 2016; Weissgerber et al., 2016).

The current research suggests mixed support for these claims. In terms of speed and accuracy of interpretation, minimalist and embellished data visualization approaches were found to have similar performance, in opposition to claims made by proponents of minimalist approaches. However, minimalist and embellished data visualization approaches also had similar levels of engagement, in opposition of claims made by proponents of embellished approaches.

Furthermore, minimalist visualizations outperformed embellished graphs in terms of memorability. However, in terms of transparency, interactive data presentations were more effective at highlighting weak findings to audiences than minimalist approaches. Thus, this research did not find benefits to using embellished approaches to data visualization, and there were unique benefits to the remaining two data presentation conventions.

Another contribution of this research to data visualization theory is that audience characteristics are an important consideration when assessing data visualization experiences and outcomes. Padney et al. (2014) explored how elaboration likelihood shaped data visualization preferences in terms of charts versus graphs. This research builds on these findings, demonstrating that motivation to elaborate also plays a significant role in participants' speed and accuracy of interpretation and recall as well as their engagement and satisfaction with the visualization.

Finally, the current research broadens the criteria by which data visualizations are judged by including broader outcomes as considerations for effective visualizations within an evaluation context. In alignment with Evaluation Influence frameworks, evaluation products represent persuasive communication meant to influence stakeholders towards social betterment (Kirkhart, 2000; Mark & Henry, 2004; Brown & Newman, 1982). Thus, the visualizations used in evaluation contexts are intended to impact stakeholders' general influence, attitudinal and cognitive, motivational, and behavioral mechanisms – and to do so in an ethical manner that accurately reflects the evaluation findings. As a result, rather than considering the speed and accuracy of interpretation as criteria for assessing the effectiveness of visualization, the current research posits that the extent to which audiences elaborate on the presented information – meaning deeply process the visualized information in a manner that considers its strength and

synthesizes it with previous knowledge – is a more critical consideration within an evaluation context.

**Elaboration Likelihood Model.** Lastly, findings from the current research have implications for the Elaboration Likelihood Model (ELM), providing insights into how principles of ELM apply within an evaluation context.

Previous research on the ELM indicates that personal involvement increases elaboration likelihood (Petty & Cacioppo, 1986; Wang, Wang, & Farn, 2009). The current research replicates these findings: interactivity in evaluation products, which increased the audiences' involvement, resulted in greater elaboration on findings than static presentations of evaluation findings. The current study also replicated measures of elaboration within an evaluation context – those shown interactive presentations of findings differentiated their attitudes based on the strength of the evidence (e.g., Petty & Cacioppo, 1986; Wang, Wang, & Farn, 2009). This finding suggests that at least some principles of ELM are relevant to the evaluation context.

However, not all findings aligned with previous research on the ELM. For example, based on the ELM, those who are low in motivation to elaborate tend to retain their original attitudes regardless of the evidence presented (e.g., Johnson & Eagly, 1989; Petty & Cacioppo, 1986; Langille et al., 2011). However, in the current study, those with low motivation to elaborate tended to show the greatest amount of attitude change, regardless of the strength of evaluation findings or how the findings were presented. The ELM poses several possibilities that may explain this unusual finding. One possibility that may explain this finding is that the evaluation or the independent researchers who conducted it are serving as peripheral clues ("experts claim this is a good program so it must be true"). Another possibility is that the presence of an evaluation can serve as an additional persuasive argument ("the organization was

willing to undergo an evaluation so it must be committed to transparency and impact"). Finally, it is possible that the evaluation could have served as a biasing factor, prompting individuals to only consider arguments in support of the evaluand (Petty & Cacioppo, 1986). The latter possibility may have some evidence, given participants' tendency to reduce their attitudes once they were presented a negative message about program costs, suggesting they had not considered these aspects of the program when reviewing the evaluation findings. However, this effect was not limited to those low in motivation to elaborate. Further research is needed to understand the role evaluation plays within elaboration processes.

An additional area where findings do not align with previous research on ELM is that elaboration did not result in stronger attitudes. These findings are surprising considering previous research findings which indicate that stronger attitudes are formed through central route elaboration processes (e.g., Crano, 1995; Haugtvedt, Schumann, Schneier, and Warren, 1994; Petty & Cacioppo, 1986; Verplanken, 1991). It is possible that evaluation findings are processed differently from logic-based findings typically used in ELM research. There are two basic structural models for persuasive communication: the syllogistic model and jurisprudence model (Areni, 2003). At the core, program evaluation assigns a value-judgement based on systematically gathered evidence (Shadish, Cook, & Leviton, 1991), which is reflective of the structure of a persuasive message outlined by Areni (2003) - the stance (the value-judgement or recommendation) which is supported by arguments (the systematically gathered evidence). However, stakeholders may not view evaluation findings as persuasive communication, and may therefore process it differently, impacting the extent to which they develop strong attitudes that persist and are resilient to new information. Further research is needed to understand how evaluation contexts impact central route processing.

#### **Future research**

The current research findings suggest several directions for future research. In terms of exploring evaluation influence, four primary areas are suggested: (1) exploring how other evaluation outputs and approaches can be designed to support evaluation influence, (2) exploring interpersonal and collective mechanisms of evaluation influence, (3) exploring evaluation influence with additional stakeholder groups, and (4) documenting current evaluation influence trends.

The current research examined how design of one evaluation output – data presentation – can impact evaluation influence. However, further research is needed to determine if other evaluation outputs and approaches can support influence mechanisms. For example, participatory approaches to evaluation may be of promise. Building on this research, participatory approaches to evaluation increase stakeholders' involvement in the evaluation. If trends from data presentation replicate to evaluation approaches, it is likely that participatory approaches to evaluation may support elaboration and evaluation influence. Similarly, other evaluation practices which are aligned with principles of the Elaboration Likelihood Model may produce promising results. For example, dedicated time for sense making of evaluation findings with stakeholders removes distractions, a key factor in increasing ability to elaborate (Kohyama & Fujihara, 1992). Combining the ELM with Mark & Henry's (2004) Evaluation Influence framework provides a theoretical grounding for how such practices increase evaluation influence framework provides a theoretical grounding for how such practices increase evaluation influence, which can be leveraged by future researchers to empirically examine the proposed links.

In addition, the current study provides empirical evidence for a general influence mechanism (elaboration), attitudinal mechanisms (attitude valence and strength), and behavioral

mechanism (donation). However, this research is limited to individual-level mechanisms. Future research should explore interpersonal and collective levels of evaluation influence. For example, to explore interpersonal behaviors, future studies could examine whether a stakeholder would be willing to advocate for legislation in support of an evaluand based on or informed by evaluation findings.

The current research is also limited to members of the public. As noted in the limitations section, other stakeholders may have unique characteristics that result in different experiences of evaluation products and levels of evaluation influence. In particular, program managers and other decision-makers represent a group of stakeholders that will likely have high levels of involvement as they have commitments to the evaluand. Future research should explore whether interactive data presentations also trigger elaboration and evaluation influence pathways for these groups.

Finally, further research is needed about current trends in the impacts of evaluations. Fleischer and Christie (2009) found limited evaluation use reported by evaluators. However, future research may find additional impacts reported by shifting the focus to evaluation influence and its mechanisms. Additionally, evaluators may not be aware of influence achieved through interpersonal and communal mechanisms, and further follow-up may be needed with the consumers of evaluation findings.

#### **Practical Implications**

The current study findings may also have several implications for evaluation practitioners. The following section outlines potential implications to evaluation product design, considerations of stakeholder characteristics, and implications of influence pathways.

In terms of evaluation product design, findings suggest that evaluators should carefully consider the design of their evaluation outputs to support evaluation influence. Based on findings, one way to do this is to seek out opportunities to design products that increase audience involvement to support elaboration processes. Additionally, while interactive data presentations are not common in evaluation practice, research findings that demonstrate differential impact of evaluation products may convince commissioners of evaluations to move beyond static reports.

Additionally, evaluation practitioners may find it beneficial to consider stakeholder's levels of motivation to elaborate. For example, stakeholders for whom the evaluation is highly relevant may have lower accuracy of interpretation and poor recall. Incorporating practices that support accurate understanding of the findings and provide reminders of the findings may be helpful for these groups. Additionally, individuals who are low in motivation to elaborate may represent an audience that could be particularly responsive to evaluation. For example, publicly funded programs that serve vulnerable populations may need support from voters that would not directly benefit from such programs and are likely to have low motivation to elaborate. Based on current study findings, these individuals' attitudes may be particularly informed by evaluation. However, additional research is needed to understand how to support these individuals in considering the strength of the evaluation findings in their assessments.

Finally, evaluation practitioners may benefit from considering their evaluations' influence through the lens of evaluation pathways for two reasons: (1) this may facilitate tracking of the impact their evaluations make and (2) this may open options to further support the influence of their evaluations if a pathway comes short of needed impact to support social betterment.

#### Conclusions

In sum, to support achieving social betterment through impactful evaluations, promoting evaluation influence mechanisms is critical. In turn, the design of evaluation outputs can support these mechanisms. The current research demonstrates that different design conventions can affect elaboration processes and can differ in their ability to facilitate influence pathways. Additionally, stakeholder characteristics such as their level of motivation to elaborate further impact evaluation influence processes. Therefore, understanding the elaboration context within evaluation can support evaluation influence towards social betterment.

More broadly, the current research adds to the evaluation literature by demonstrating how a general theory of cognition and behavior can be translated to evaluation settings. Henry and Mark (2003) observe that the lack of research on evaluation (ROE) has resulted in both practice and debate on key issues within the field not being based in "rigorous, systematic evidence" but instead on "personal experience, interactions, reading, training, or intuition" (Henry & Mark, 2003, p. 69). In response, they provide an extensive research agenda to guide future ROE work. However, research gaps in evaluation will not be solved overnight, and evaluation practitioners need guidance in the meantime. In the absence of research conducted specifically in the evaluation context, sister fields such as psychology and sociology can provide meaningful insights through the translation of general theories of human behavior and cognition. These theories represent current best knowledge based on rigorous research and need to be empirically tested within the evaluation context.

Overall, the Petty and Cacioppo's (1986) Elaboration Likelihood Model provides promising insights which may serve as a beneficial link between evaluation practices and products and evaluation influence. Deepening the understanding of how evaluation product

designs effect stakeholder processing of evaluation findings and their resulting reactions can be a powerful tool towards promoting evaluation influence.

#### References

Agresti, A. (2007). An introduction to categorical data analysis (2nd ed.). Hoboken, NJ: Wiley.

- Ajzen, I. (1991). The Theory of Planned Behavior. Organizational Behavior and Human Decision Processes, 50, 179 – 211.
- 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.
   DOI: 10.1177/1098214017717015
- Areni, C. S. (2003). Elaboration Likelihood Model perspective. *Psychology and Marketing*, 20(4), 349 – 375. DOI: 10.1002/mar.10077
- Azzam, T. & Whyte, C. E. (2018). Evaluative feedback delivery and the factors that affect success. *Evaluation and Program Planning*, 67, 148 159. DOI: 10.1016/j.evalprogplan.2017.11.005
- Azzam, T., Evergreen, S., Germuth, A. A., & Kistler, S. J. (2013). Data visualization and evaluation. In T. Azzam & S. Evergreen (Eds.), *Data visualization, part 1. New Directions for Evaluation*, 139, 7-32. DOI: 10.1002/ev.20065
- Bateman, S., Mandryk, R. L., Gutwin, C., Genest, A., McDine, D., & Brooks, C. (2010). Useful junk? The effects of visual embellishment on comprehension and memorability of charts, *CHI 2010*, Atlanta, GA, April 10–15, 2010. New York, NY: ACM.
- Bechar, S. & Mero-Jaffe, I. (2014). Who is afraid of evaluation? Ethics in evaluation research as a way to cope with excessive evaluation anxiety: Insights from a case study. *American Journal of Evaluation*, 35, 364-376. doi: 10.1177/1098214013512555
- Blueshtein, M. (2018). Guilt and pro-social behavior amongst workers in an online labor market. Journal of Business & Economic Policy, 5(4), 1 - 5.

- Bray, J. H., & Maxwell, S. E. (1985). *Multivariate analysis of variance*. Thousand Oaks, CA: Sage.
- Brown, R. D. & Newman, D. L. (1982). An investigation of the effect of different data presentation formats and order of arguments in a simulated adversary evaluation. *Educational Evaluation and Policy Analysis*, 14 (2), 197-203. DOI: 10.3102/01623737004002197
- Bryne, L., Angus, D., & Wiles, J. (2016). Acquired codes of meaning in data visualization and infographics: Beyond perceptual primitives. *IEEE Transaction on Visualization and Computer Graphics*, 22 (1), 509 – 518. DOI: 10.1109/TVCG.2015.2467321
- Brynjolfosson, E., Horton, J. J., Ozimek, A., Rock, D., Sharma, G., & TuYe, H. (2020). COVID-19 and remote work: An early look at US data. *National Bureau of Economics Working Paper Series: Working Paper 27344*. Cambridge, MA: National Bureau of Economic Research.
- Bryson, J. M., Patton, M. Q., & Bowman, R. A. (2011). Working with evaluation stakeholders:
  A rationale, step-wise approach and toolkit. *Evaluation and Program Planning*, 34, 1-12.
  DOI: 10.1016/j.evalprogplan.2010.07.001
- Cacioppo, J. T., Petty, R. E., & Morris, K. (1983). Effects of need for cognition on message evaluation, recall, and persuasion. *Journal of Personality and Social Psychology*, 45, 805-818.
- Christenson, D. P. & Glick, D. M. (2013). Crowdsourcing panel studies and real-time experiments in MTurk. *The Political Methodologist*, 20 (2), 27 32.
- Cleveland, W. S. & McGill, R. (1985). Graphical perception and graphical methods for analyzing scientific data. *Science*, 229(4716), 828 833.

- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York: Psychology Press.
- Cook, R. D., & Weisberg, S. (1982). *Residuals and influence in regression*. New York: Chapman & Hall.
- Crano, W. D. (1995). Attitude strength and vested interest. In R. E. Petty & J. A. Krosnick (Eds.), *Attitude strength antecedents and consequences* (pp. 131 158). New York, NY: Psychology Press.
- Crano, W. D., Brewer, M. B., & Lac, A. (2015). *Principles and methods of social research* (3<sup>rd</sup> Ed.). New York, NY: Routledge.
- Crano, W. D. & Sivacek, J. (1982). Social reinforcement, self-attribution, and the foot-in-thedoor phenomenon. *Social Cognition*, 1(12), 110 – 125. DOI: 10.1521/soco.1982.1.2.110
- Difallah, D., Filatova, E., & Ipeirotis, P. (2018). *Demographics and dynamics of Mechanical Turk workers*. Paper presented at WSDM 2018: The Eleventh ACM International
  Conference on Web Search and Data Mining, Marina Del Rey, CA, February 5-9, 2018.
  DOI: 10.1145/3159652.3159661
- Ditto, P. H. & Lopez, D. F. (1992). Motivated skepticism: Use of differential decision criteria for preferred and nonpreferred conclusions. *Journal of Personality and Social Psychology*, 63(4), 568 584. DOI: 10.1037//0022-3514.63.4.568
- Donaldson, S. I., Gooler, L. E., & Scriven, M. (2002). Strategies for managing evaluation anxiety: Toward a psychology of program evaluation. *American Journal of Evaluation*, 23, 261–273. doi:10.1016/S1098-2140(02)00209-6

- Evergreen, S., & Metzner, C. (2013). Design principles for data visualization in evaluation. In T. Azzam & S. Evergreen (Eds.), *Data visualization, part 2. New Directions for Evaluation*, 140, 5 – 20.
- Fisher, D. (2010). Animation for visualization: Opportunities and drawbacks. In J. Steele and N.
  Iliinsky (Eds.), *Beautiful visualization: Looking at data through the eyes of experts*(pp.329-352). Sebastopol, CA: O'Reilly Media.
- Fleischer, D. N. & Christie, C. A. (2009). Evaluation use: Results from a survey of U.S.
  American Evaluation Association members. *American Journal of Evaluation*, 30(2), 158-175. DOI: 10.1177/1098214008331009
- Fleming, M. A. (2011). Attitudes, persuasion, and social influence: Applying social psychology to increase evaluation use. In M. M. Mark, S. I. Donaldson, & B. Campbell (Eds.), *Social psychology and evaluation* (pp. 212 – 243). Gilbert Press.
- Gargani, J. & Donaldson, S. I. (2011). What works for whom, where, why, for what, and when?
  Using evaluation evidence to take action in local contexts. In H. T. Chen, S. I.
  Donaldson, & M. M. Marks (Eds.), *Advancing validity in outcome evaluation: Theory and practice. New Directions for Evaluation*, 130, 17-30, DOI: 10.1002/ev.362
- Gillian, D. J. & Richman, E. H. (1994). Minimalism and the syntax of graphs. *The Journal of Human Factors and Ergonomics Society*, 36(4), 619–644.
- Haugtvedt, C. P., Schumann, D. W., Schneier, W. L., Warren, W. L. (1994). Advertising repetition and variation strategies: Implications for understanding attitude strength. *Journal of Consumer Research*, 21, 176 – 189.

- Heer, J., & Bostock, M. (2010). Crowdsourcing graphical perception: Using Mechanical Turk to assess visualization design. Paper presented at CHI 2010, April 10 -15, 2010, Atlanta, GA.
- Henry, G. T. (2000). Why not use? in V. Caracelli and H. Preskill (Eds.) The Expanding Scope of Evaluation Use. *New Directions for Evaluation*, 88, pp. 85-98. San Francisco, CA: Jossey-Bass. DOI: 10.1002/ev.1193
- Henry, G. T. & Mark, M. M. (2003). Beyond use: Understanding evaluation's influence on attitudes and actions. *American Journal of Evaluation*, 24(3), 293 314.
- Herbert, J. L. (2014). Researching evaluation influence: A review of the literature. *Evaluation Review*, 38(5), 388-419. DOI: 10.1177/0193841X14547230
- Hung, Y. & Parsons, P. (2017). Assessing user engagement in information visualization. In CHI EA'17: Extended Abstracts on Human Factors in Computing Systems, pp. 1708-1717, New York, New York, USA
- Johnson, B. T. & Eagly, A. H. (1989). Effects of involvement on persuasion: A meta-analysis. *Psychological Bulletin*, 106(2), 290 – 314.
- Kao, D. T. (2011). Message sidedness in advertisement: The moderating roles of need for cognition and time pressure in persuasion. *Scandinavian Journal of Psychology*, 52, 329 340. DOI: 10.1111/j.1467-9450.2011.00882.x
- Kohyama, T. & Fujihara, T. (1992). Effects of need for cognition and distraction on issuerelevant elaboration and attitude change. *Experimental Social Psychology Research*, 31(3), 222 – 230. DOI: 10.2130/jjesp.31.222

- Kirkhart, K. (2000). Reconceptualizing Evaluation Use: An Integrated Theory of Influence, in V.
   Caracelli and H. Preskill (Eds.) The Expanding Scope of Evaluation Use. *New Directions for Evaluation*, 88, pp. 5–24. San Francisco, CA: Jossey-Bass. DOI: 10.1002/ev.1188
- Langille, J. D., Berry, T. R., Reade, I. L., Witcher, C. Loitz, C. C., & Rodgers, W. M. (2011). Strength of messaging in changing attitudes in a workplace wellness program. *Health Promotion Practice*, 12(2), 303 – 311. DOI: 10.1177/1524839909336650
- Lien, N. (2001). Elaboration Likelihood Model in consumer research: A review. *Proc. Natl. Sci. Counc.*, 11(4), 301 – 311.
- Malaviya, P. (2007). The moderating influence of advertising context on ad repetition effects:
  The role of amount and type of elaboration. *Journal of Consumer Research*, 34, 32 40.
  DOI: 10.1086/513044
- Mark, M. M. & Henry, G. T. (2004). The mechanisms and outcomes of evaluation influence. *Evaluation*, 10(1), 35-37. DOI: 10.1177/1356389004042326
- Mason, S., & Azzam, T. (2018). 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. DOI: 10.1177/1098214018778808
- Natter, H. M. & Berry, D. C. (2005). Effects of active information processing on the understanding of risk information. *Applied Cognitive Psychology*, 19, 123–35. DOI: 10.1002/acp.1068.
- Newman, D. L., Brown, R. D., & Braskamp, L. A. (1980). Communication theory and the utilization of evaluation. In L. A. Braskamp & R. D. Brown (Eds.), *Utilization of Evaluation Information, New Directions for Evaluation*, 1980, pp. 29 – 35. San Francisco, CA: Jossey-Bass. DOI: 10.1002/ev.1234

- Oliver, M. L. (2008). Evaluation for emergency response: Humanitarian aid agencies and evaluation use. Retrieved from Georgia Institute of Technology Library. URI: http://hdl.handle.net/1853/24629
- Olson, C. L. (1976). On choosing a test statistic in multivariate analysis of variance. *Psychological Bulletin*, 83(4), 579-586.
- Otten, J. J., Cheng, K., Drewnowski, A. (2015). Infographics and public policy: Using data visualization to convey complex information. Health Affairs, 34(11), 1901-1907
- Pandey, A.V., Manivannan, A., Nov, O., Satterthwaite, M. L., Bertini, E. (2014). The persuasive power of data visualization. *New York University Public Law and Legal Theory Working Papers*, Paper 474.
- Patton, M. Q. (1998). Discovering process use. *Evaluation*, 4(2), 225-233. DOI: 10.1177/13563899822208437
- Patton, M. Q. (2000). Overview: Language matters. *New Directions for Evaluation*, 2000(86), 5-16. DOI: 10.1002/ev.1168
- Preskill, H. & Catsambas, T. T. (2006). *Reframing evaluation through appreciative inquiry*. Thousand Oaks, CA: Sage Publications, Inc.
- Petty, R. E., & Cacioppo, J. T. (1986). The Elaboration Likelihood Model of Persuasion. Advances in Experimental Social Psychology, 19, 123 – 205.
- Petty, R. E., Cacioppo, J. T., & Heesacker, M. (1985). Persistence of persuasion: A test of the Elaboration Likelihood Model. Unpublished manuscript. University of Missouri, Columbia.

- Petty, R. E., Fabrigar, L. R. & Wegener, D. T. (2003). Emotional factors in attitudes and persuasion. In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith (Eds.), *Handbook of* affective sciences, pp. 752 – 777. Oxford: Oxford University Press.
- Petty, R. E., Wells, G. L., & Brock. T. C. (1976). Distraction can enhance or reduce yielding to propaganda: Thought disruption versus effort justification. *Journal of Personality and Social Psychology*, 34, 874 – 884.
- Schonlau, M., & Peters, E. (2008). Graph comprehension: An experiment in displaying data as bar charts, pie charts and tables with and without gratuitous 3<sup>rd</sup> dimension, RAND Working Paper, WR-618.
- Schumann, D. W., Kotowski, M. R., Ahn, H. Y. A., Haugtvedt, C. P. (2012). The Elaboration Likelihood Model 30 Year Review. In S. Rogers & E. Thorson (Eds.), *Advertising Theory* (pp. 51-68). New York, NY: Routledge.
- Shadish, W. R., Cook, T. D., Leviton, L. C. (1991). *Foundations of program evaluation: Theories of practice*. Newbury Park, CA: Sage Publications.
- Sherif, M. & Hovland, C. I. (1961). Social judgement: Assimilation and contrast effects in communication and attitude change. Oxford, England: University Press.
- Sherif, C. W., Kelly, M., Rodgers, H. L., Sarup, G., & Tittler, B. I. (1973). Personal involvement, social judgement, and action. *Journal of Personality and Social Psychology*, 27(3), 311 – 328.
- Shulha, L. & Cousins, B. J. (1997). Evaluation use: Theory, research, and practice since 1986. *American Journal of Evaluation*, 18 (3), 195-208. DOI: 10.1177/109821409701800302
- Siegrist, M. (1996). The use or misuse of three-dimensional graphs to represent lowerdimensional data. *Behaviour and Information Technology*, 15(2), 96-100.

- Tormala, Z. L., Briñol, P., & Petty, R. E. (2007). Multiple roles for source credibility under high elaboration: It's all in the timing. *Social Cognition*, 25(4), 536 552.
- Tufte, E. R. (1983). *The Visual Display of Quantitative Information*. Cheshire, Connecticut: Graphics Press.
- Verplanken, B. (1991). Persuasive communication of risk information: A test of cue versus message processing effects in a field experiment. *Personality and Social Psychology Bulletin*, 17(2), 188 – 193. DOI: 10.1177/014616729101700211
- Wang, K., Wang, E. T. G., Farn, C. K. (2009). Influence of web advertising strategies, consumer goal-directedness, and consumer involvement on web advertising effectiveness. *International Journal of Electronic Commerce*, 13(4), 67 95. DOI: 10.2753/JEC1086-441513040
- Weinfurt, K. P. (1995). Multivariate analysis of variance. In Grimm, L. G. & P. R. Yarnold (Eds.), *Reading and understanding multivariate statistics* (pp. 245-276). Washington, DC: American Psychological Association.
- Weiss, C. H. (1998). Have we learned anything new about the use of evaluation? *American Journal of Evaluation*, 19 (1), 21 33. DOI: 10.1177/109821409801900103
- Weissgerber, T. L., Garovic, V. D., Savic, M., Winham, S. J., Milic, N. M. (2016). From static to interactive: Transforming data visualization to improve transparency. *PLoS Biol*, 14(6), 1 8.
- 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 Biolchemistry*, 292(50), 20592 20598. DOI: 10.1074/jbc.RA117.000147
- Yarborough, D. B., Shulha, L. M., Hopson, R. K., & Carruthers, F. A. (2011). *The program evaluation standards* (3rd ed.). Thousand Oaks, CA: Sage.
- Zanna, M. P., & Cooper, J. (1974). Dissonance and the pill: an attribution approach to studying the arousal properties of dissonance. *Journal of personality and social psychology*, 29(5), 703.

# Appendix A. Previous Operationalizations of Evaluation Influence

Level	Process	Definition				
	Direction/valence	A change in attitude, positive or negative, about the				
Individual	of attitude	intervention being evaluated or about the disaster relief				
	change	program as a whole.				
		A change in how an individual acts as a direct result of				
Individual	Behavior change	participating in an evaluation or reading an evaluation				
		report.				
Individual	Salience	The importance an individual gives to an idea.				
Individual	Elaboration	The extent to which a person thinks about or mentally				
		processes a given issue.				
Individual	Priming	Bringing a given idea or concept to the forefront, setting				
		it up to have an impact on judgements or decisions.				
Individual	Skill Acquisition	Increasing one's competence in a skill area via				
	1	participating in the process of evaluation.				
Interpersonal	Justification	Using the conclusions from an evaluation report to back				
•	v	up one's prior convictions about an issue.				
Interpersonal	Persuasion	Attitude change that one individual attempts to bring				
-		Apoul III allotter.				
Interpersonal	Change agent	change				
-		Change in agreed upon principles about how to conduct				
Interpersonal	Social norms	oneself in a given setting				
	Minority-opinion	Altering the opinion of those whose attitude does not				
Interpersonal	influence	align with the majority				
Collective	inginence	Getting an issue on the docket for public and/or				
Action	Agenda setting	government consideration				
Collective	Policy-oriented					
Action	learning	Attitude change about policy objectives.				
Collective		Actual shift in internal policy in the operating environment.				
Action	Policy change					
Collective		External ration shift beyond the anomating environments				
Action	Dijjusion	External policy shift beyond the operating environments.				

Table A1. Oliver's (2008) definitions of evaluation influence processes.

#### **Appendix B. Recruitment and Debrief Texts**

#### A1. Scenario Pilot Recruitment Text:

Title: Short Survey on Interest in Non-profits (~5 min)

**Description:** This is a research project about how interested people would be in reading more information and thinking more about a non-profit program based on its description. If you decide to participate, you will read brief descriptions of non-profit programs and complete survey questions asking about your interest in the program. Task is anticipated to take 8 minutes and you will be compensated \$0.50 for survey completion.

#### A2. Scenario Pilot Debrief Text:

Thank you for participating in this study! There is more to this study than what I've told you so far. Sometimes, in psychological research it is necessary to not tell people everything about the study at the beginning. If we did, it might affect how they respond to the questions asked and the tasks involved, and this would change the results in a way that would make them invalid. **The non-profit programs you read about are not real.** 

The purpose of this study is to gauge how interested people would be in reading more information and thinking more about these programs. You were asked to read short descriptions of several non-profit programs and indicate how interested you would be in the program, does it reflect your core values, and if you would be personally impacted by the program if it was offered in your community. This information will be used to refine study materials for a broader study on how to best present information about the effectiveness of non-profits and other social programs. Again, we thank you for your participation in this study. If you have any further questions or concerns, please feel free to contact me (agnieszka.rykaczewska@cgu.edu) or my research supervisor (tarek.azzam@cgu.edu) for more information. 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 <u>irb@cgu.edu</u>.

## **B1. Evidence Strength Pilot Recruitment Text:**

Title: Short Survey on Opinions on Non-profit outcomes (~2 min)

**Description:** This is a research project about what outcomes people consider effective for a nonprofit. You will read information a non-profit program and hypothetical results of a study about the program's outcomes, and will be asked to complete a 5 minute questionnaire, asking about whether you are convinced by the hypothetical results that the program is effective. Task is anticipated to take 2 minutes and you will be compensated \$0.25 for survey completion.

#### **B2. Evidence Strength Pilot Debrief Text:**

Thank you for participating in this study!

The purpose of this pilot study is to gauge how effective a program's outcomes need to be for people to find the results convincing. You read about a non-profit and indicated how convincing you found a series of hypothetical results. This information will be used to refine study materials for a broader study on how to best present information about the effectiveness of non-profits and other social programs.

Again, we thank you for your participation in this study. If you have any further questions or concerns please feel free to contact me (agnieszka.rykaczewska@cgu.edu) or my research

supervisor (tarek.azzam@cgu.edu) for more information. 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.

Please click the arrow below to receive your MTurk completion code.

## C1. Phase 1 Recruitment Text (Initial Questionnaire):

Title: Survey on Presentation Style of Non-profit Information 1.0 (~15-20 min)

**Description:** This is a research project about different approaches to present information about social and non-profit programs. If you decide to participate, you will read some information about a non-profit program, and complete to a 15 survey asking about the information you read, your opinions of the program and how the information was presented, and demographic questions. Additionally, we will follow-up with you in about a week to ask additional questions about the information you read and your opinions of the program, which will take an additional 5 minutes to complete. You will be compensated \$2.40 for completing the initial questionnaire and an additional \$1.00 for completing the follow-up questionnaire.

#### **<u>C2. Phase 1 Memorability Follow-up Recruitment Text:</u>**

Message from Agnieszka Rykaczewska (agnieszka.rykaczewska@cgu.edu)

-----

Greetings!

About a week ago, you completed a survey on Amazon MTurk about the WISH program. This is a brief follow-up survey.

If you decide to participate, the survey will ask additional questions about the information you

## C3. Phase 1 Debrief

Thank you for your participation in this research project.

There is more to this study than what I've told you so far. Sometimes, in psychological research it is necessary to not tell people everything about the study at the beginning. If we did, it might affect how they respond to the questions asked and the tasks involved, and this would change the results in a way that would make them invalid. **The non-profit program and the study findings that you read were not real. The WISH program does not exist.** 

The goal of this study was to test whether different data visualization approaches would result in different speed and accuracy of data interpretation and influence attitudes and memory differently. You were asked to read a description of a non-profit program and presented with study findings about that program's effectiveness. You were randomly assigned to presented either traditional minimalist graphs or embellished cartoonish graphs. We then measured how

this affected your opinions of the non-profit program and whether you remembered the study findings a week later. If you knew that this non-profit program and the study findings were not real, it may have shaped your opinions and memory.

Again, we thank you for your participation in this study. The findings of this study will help us better communicate data and research findings about the effectiveness of non-profit and other social programs. If you have any further questions or concerns please feel free to contact me (agnieszka.rykaczewska@cgu.edu) or my research supervisor (tarek.azzam@cgu.edu) for more information. 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.

## D1. Phase 2 Recruitment Text (Initial Survey)

Title: Survey on Presentation Style of Non-profit Information 2.0 (~15-20 min)

**Description**: This is a study ONLY for workers who DID NOT complete the first wave of this research. This is a research project about different approaches to present information about social and non-profit programs. If you decide to participate, you will read some information about a non-profit program, and complete to a 15 to 20-minute survey asking your opinions of the program and how the information was presented, your willingness to support the program, and demographic questions. Additionally, we will follow-up with you in about a week to ask additional questions about the information you read and your opinions of the program, which will take an additional 10 minutes to complete. You will be compensated \$2.40 for completing the initial questionnaire and an additional \$1.60 for completing the follow-up questionnaire.

# **D2. Phase 2 Persistence Follow-up Recruitment Text:**

Message from Agnieszka Rykaczewska (agnieszka.rykaczewska@cgu.edu)

-----

Greetings!

About a week ago, you completed a survey on Amazon Mturk about the WISH program. This is a brief follow-up survey.

If you decide to participate, the survey will ask additional questions about the information you read and your opinions of the program.

The follow-up survey will take approximately 5 minutes to complete, and you will be

compensated \$1.60 for completing it.

To complete the follow-up survey, please click the following link: [Survey Link]

Please let me know if you have any questions or concerns. Thank you for your support of this research.

\_\_\_\_\_

Greetings from Amazon Mechanical Turk,

The message above was sent by an Amazon Mechanical Turk user.

Please review the message and respond to it as you see fit.

Sincerely,

Amazon Mechanical Turk

https://www.mturk.com

-----

#### **D3.** Phase 2 Resilience Task Recruitment Text

This is the second part of the questionnaire, which is optional and should take about 5 minutes. If you complete this part of the questionnaire, you will be compensated an additional \$1.60. If you do not wish to complete this part of the questionnaire, simply skip the questions and go to the end to receive your submission code.

#### **D4. Phase 2 Debrief Text**

Thank you for your participation in this research project.

There is more to this study than what I've told you so far. Sometimes, in psychological research it is necessary to not tell people everything about the study at the beginning. If we did, it might affect how they respond to the questions asked and the tasks involved, and this would change the results in a way that would make them invalid. **The non-profit program and the study findings that you read were not real. The WISH program does not exist. You will receive your full earnings from this HIT, regardless of what you marked as your donation choice in the first part of this study.** 

The goal of this study was to test whether different data visualization approaches would influence attitudes and behaviors differently. You were asked to read a description of a nonprofit program and presented with study findings about that program's effectiveness. You were randomly assigned to presented either traditional minimalist graphs or embellished cartoonish graphs. We then measured how this affected your opinions of the non-profit program and whether you would be willing to engage in different behaviors that would be supportive of the program, including donating your HIT earnings to the program. Some individuals were randomly assigned to be contacted a week later to test whether your opinions of the program would change

over time. Others were randomly assigned to be presented with a negative message about the non-profit to see whether your opinion would change based on this information. **If you knew that this non-profit program and the study findings were not real, it may have shaped your opinions and behaviors. Again, you will receive your full earnings from this HIT, regardless of any answers you provided to the questions in this survey.** 

We thank you for your participation in this study. The findings of this study will help us better communicate data and research findings about the effectiveness of non-profit and other social programs. If you have any further questions or concerns please feel free to contact me (agnieszka.rykaczewska@cgu.edu) or my research supervisor (tarek.azzam@cgu.edu) for more information. 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.

#### E1. Phase 3 Recruitment Text (Initial Survey)

**Title:** Survey on Presentation Style of Non-profit Information 3.0 (~20-25 min) **Description:** This is a study ONLY for workers who DID NOT complete the first wave of this research. This is a research project about different approaches to present information about social and non-profit programs. If you decide to participate, you will read some information about a non-profit program, and complete to a 15 to 20-minute survey asking your opinions of the program and how the information was presented, your willingness to support the program, and demographic questions. Additionally, in about a week, we will send a follow-up survey that asks additional questions about the information you read and your opinions of the program, which will take an additional 5 minutes to complete. You will be compensated \$2.40 for completing the initial questionnaire and an additional \$1.60 bonus for completing follow-up questionnaire.

# E2. Phase 3 Persistence Follow-up Recruitment Text:

Message from Agnieszka Rykaczewska (agnieszka.rykaczewska@cgu.edu)

-----

Greetings!

About a week ago, you completed a survey on Amazon Mturk about the WISH program. This is a brief follow-up survey.

If you decide to participate, the survey will ask additional questions about the information you read and your opinions of the program.

The follow-up survey will take approximately 5 minutes to complete, and you will be

compensated \$1.60 for completing it.

To complete the follow-up survey, please click the following link: [Survey Link]

-----

Greetings from Amazon Mechanical Turk,

The message above was sent by an Amazon Mechanical Turk user.

Please review the message and respond to it as you see fit.

Sincerely,

Amazon Mechanical Turk

https://www.mturk.com

#### E3. Phase 3 Resilience Task Recruitment Text

This is the second part of the questionnaire, which is optional and should take about 5 minutes. If you complete this part of the questionnaire, you will be compensated an additional \$1.60. If you do not wish to complete this part of the questionnaire, simply skip the questions and go to the end to receive your submission code.

#### E4. Phase 3 Debrief

Thank you for your participation in this research project.

There is more to this study than what I've told you so far. Sometimes, in psychological research it is necessary to not tell people everything about the study at the beginning. If we did, it might affect how they respond to the questions asked and the tasks involved, and this would change the results in a way that would make them invalid. **The non-profit program and the study findings that you read were not real. The WISH program does not exist. You will receive your full earnings from this HIT, regardless of what you marked as your donation choice in the first part of this study.** 

The goal of this study was to test whether different data visualization approaches would influence attitudes and behaviors differently. You were asked to read a description of a nonprofit program and presented with study findings about that program's effectiveness. You were randomly assigned to presented either traditional minimalist graphs or embellished cartoonish graphs. We then measured how this affected your opinions of the non-profit program and whether you would be willing to engage in different behaviors that would be supportive of the program, including donating your HIT earnings to the program. Some individuals were randomly assigned to be contacted a week later to test whether your opinions of the program would change

over time. Others were randomly assigned to be presented with a negative message about the non-profit to see whether your opinion would change based on this information. **If you knew that this non-profit program and the study findings were not real, it may have shaped your opinions and behaviors. Again, you will receive your full earnings from this HIT, regardless of any answers you provided to the questions in this survey.** 

We thank you for your participation in this study. The findings of this study will help us better communicate data and research findings about the effectiveness of non-profit and other social programs. If you have any further questions or concerns please feel free to contact me (agnieszka.rykaczewska@cgu.edu) or my research supervisor (tarek.azzam@cgu.edu) for more information. 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.

## **Appendix C. Scenario Text**

WISH is a non-profit program that believes women should have flexibility in their careers and connects women with opportunities to work from home. WISH's latest initiative advocates for more remote work opportunities for women and connects women with companies that offer remote work positions with flexible hours. Additionally, WISH works with local non-profits to provide career training programs geared towards remote work. WISH claims that the program empowers women through improved economic opportunity and better work-life balance. Recently, independent researchers investigated how effective WISH was in promoting remote work opportunities for women.

# Appendix D. Motivation to Elaborate Scale

1.	How interested	d are y	ou in le	learning about the effectiveness of this program?				s of this program?	
	Not at all -1	2	3	4	5	6	7-Ve	ery interested	
2.	How much does this program relate to your core values?								
	Not at all -1	2	3	4	5	6	7-Ve	ery related to core values	
3.	3. To what extent would offering this program in your community have practical implications for your personal life?								
	No implication	ns -1	2	3	4	5	6	7-Significant implications	
Scale ]	Internal Consist	tency A	Analyse	<u>es</u>					

Based on Study 1 responses, the Motivation to Elaborate Scale showed high internal consistency

(Cronbach's alpha = 0.813, items = 3, N = 202).

## **Appendix E. Attitude Scales**

## **Pre-test Attitudes Scales**

I would like to learn your opinions about programs that address economic opportunity and work-life balance for women. The adjectives below may or may not represent your opinion. For each adjective, **please check the line that best expresses your opinion about programs that address economic opportunity and work-life balance for women.** 

Bad\_\_\_\_\_\_\_\_\_GoodUnimportant\_\_\_\_\_\_\_\_\_\_ImportantImpractical\_\_\_\_\_\_\_\_\_\_PracticalWorthless\_\_\_\_\_\_\_\_\_\_YorthwhileUnnecessary\_\_\_\_\_\_\_\_\_\_Necessary

The following questions ask your opinions about the **WISH program**. Please indicate the extent to which you disagree or agree with the following statements.

1. Providing the WISH program is a good use of the non-profit's resources.

Strongly disagree Disagree Slightly disagree Neutral Slightly agree Agree Strongly agree

The WISH program is an important service for women in my community.
 Strongly disagree Disagree Slightly disagree Neutral Slightly agree Agree Strongly agree

Participating in the WISH program is a waste of a woman's time.<sup>5</sup>
 Strongly disagree Disagree Slightly disagree Neutral Slightly agree Agree Strongly agree

4. I worry about women that participate in the WISH program.<sup>5</sup>

<sup>&</sup>lt;sup>5</sup> Item reverse coded.

Strongly disagree Disagree Slightly disagree Neutral Slightly agree Agree Strongly agree

5. I think that WISH is a good program.

Strongly disagree Disagree Slightly disagree Neutral Slightly agree Agree Strongly agree

Scale Internal Consistency Analyses

**Global Attitudes**: Based on pre-treatment responses, the five item Global Attitudes measure had high internal consistency (Cronbach's alpha = 0.916, items = 5, N = 202). Similar internal consistency was detected in post-treatment responses (Cronbach's alpha = 0.933, items = 5, N = 202).

**Evaluand-specific Attitudes**: Based on pre-treatment responses, the five item Evaluand-specific Attitudes Measure had high internal consistency (Cronbach's alpha = 0.737, items = 5, N = 202). Similar internal consistency was detected in post-treatment responses (Cronbach's alpha = 0.773, items = 5, N = 201).

## **Appendix F. Treatments**

### A. Condition 1: Minimalist/Minimalist Data Visualization with Strong Evidence

The following chart presents the results of the WISH program effectiveness study. Women were randomly assigned to either the participate in the program (Program Participants) or to continue their normal daily activities (Comparison Group).

The percent of women who reported that they were "extremely satisfied" with their work-life balance increased from 10% to 55% among program participants. The comparison group did not see a similar increase.

Percent of Women "Extremely Satisfied"



B. Condition 2: Minimalist/Minimalist Data Visualization with Weak Evidence The following chart presents the results of the of the WISH effectiveness study. Women were randomly assigned to either the participate in the program (Program Participants) or to continue their normal daily activities (Comparison Group).

The percent of women who reported that they were "extremely satisfied" with their worklife balance increased from 10% to 15% among program participants. The comparison group did not see a similar increase.



## C. Condition 3: Embellished Data Visualization with Strong Evidence

The following chart presents the results of the WISH program effectiveness study. Women were randomly assigned to either the participate in the program (Program Participants) or to continue their normal daily activities (Comparison Group).

The percent of women who reported that they were "extremely satisfied" with their work-life balance increased from 10% to 55% among program participants. The comparison group did not see a similar increase.



WISH Program

## D. Condition 4: Embellished Data Visualization with Weak Evidence

The following chart presents the results of the of the WISH effectiveness study. Women were randomly assigned to either the participate in the program (Program Participants) or to continue their normal daily activities (Comparison Group).

The percent of women who reported that they were "extremely satisfied" with their work-life balance increased from 10% to 15% among program participants. The comparison group did not see a similar increase.



Additionally, women who participated in the WISH program had a median 8% increase in their monthly income. The comparison group did not see a similar increase.



# **Appendix G. Interpretation Survey Items**

## **Condition: Strong Evidence**

1. Please describe the study findings in your own words.

- 2. What percent of women reported feeling "extremely satisfied" with their work-life balance after participating in WISH?
  - A. 10%
  - **B**. 11%
  - C. 13%
  - D. 55%
  - E. None of the above.
- 3. Participating in WISH \_\_\_\_\_\_ the number of women who reported feeling "extremely satisfied" with their work-life balance.
  - A. Increased
  - B. Decreased
  - C. Did not change
- 4. Of those who <u>did not</u> participate in the program, what percent of women reported feeling "extremely satisfied" with their work-life balance at Time 2?
  - A. 10%
  - B. 11%
  - C. 13%
  - D. 55%
  - E. None of the above.
- 5. Among those who participated in WISH, the percent of women who reported feeling "extremely satisfied" with their work-life balance increased by \_\_\_\_\_.
  - A. 2%
  - B. 3%
  - C. 45%
  - D. 55%
  - E. None of the above.

- 6. On average, program participants increased their monthly income by \_\_\_\_\_ more than women who did not participate in WISH.
  - A. \$50
  - B. \$100
  - C. \$150
  - D. \$200
  - E. None of the above.

## **Condition: Weak Evidence**

1. Please describe the study findings in your own words.

- 2. Participating in WISH \_\_\_\_\_\_ the number of women who reported feeling "extremely satisfied" with their work-life balance.
  - A. Increased
  - B. Decreased
  - C. Did not change
- 3. What percent of women reported feeling "extremely satisfied" with their work-life balance after participating in WISH?
  - A. 10%
  - **B.** 11%
  - C. 13%
  - D. 15%
  - E. None of the above.
- 4. Of those who <u>did not</u> participate in the program, what percent of women reported feeling "extremely satisfied" with their work-life balance at Time 2?
  - A. 10%
  - **B**. 11%
  - C. 13%
  - D. 15%
  - E. None of the above.

- 5. Among those who participated in WISH, the percent of women who reported feeling "extremely satisfied" with their work-life balance increased by \_\_\_\_\_.
  - A. 5%
  - **B.** 10%
  - C. 15%
  - D. 20%
  - E. None of the above.
- 6. On average, program participants increased their monthly income by \_\_\_\_\_ more than women who did not participate in WISH.
  - A. \$50
  - **B.** \$100
  - C. \$150
  - D. \$200
  - E. None of the above.

## **Appendix H. Perceptions Scales**

#### A. Phase 1 and 2 Engagement Measure:

I would like to learn about your opinion on how the study findings were presented. Overall, how much do you disagree or agree with the following statements?

While reading this chart, I found its look to be pleasing.
 Strongly disagree Disagree Slightly disagree Neutral Slightly agree Agree
 Strongly agree

2. The layout of this chart is clear and balanced.

Strongly disagree Disagree Slightly disagree Neutral Slightly agree Agree Strongly agree

3. While reading this chart, I felt absorbed to the extent that I was not aware of my surroundings.

Strongly disagree Disagree Slightly disagree Neutral Slightly agree Agree Strongly agree

While reading this chart, time seemed to pass quickly.
 Strongly disagree Disagree Slightly disagree Neutral Slightly agree Agree
 Strongly agree

While reading this chart, I enjoyed and accepted any challenges it presented.
 Strongly disagree Disagree Slightly disagree Neutral Slightly agree Agree
 Strongly agree

6. While reading this chart, I had to think carefully, deeply, or reflectively.

Strongly disagree Disagree Slightly disagree Neutral Slightly agree Agree Strongly agree

7. While reading this chart, I learned something that I had not known before (e.g., a new fact, concept, or piece of information).

Strongly disagree Disagree Slightly disagree Neutral Slightly agree Agree Strongly agree

8. I learned and figured out how to read this chart along the way.
 Strongly disagree Disagree Slightly disagree Neutral Slightly agree

Strongly agree

9. While reading this chart, I felt as though I was moving in or through it to learn about its content or message.

Agree

Strongly disagree Disagree Slightly disagree Neutral Slightly agree Agree Strongly agree

10. While reading this chart, I was exploring its features and content in a gradual fashion.Strongly disagree Disagree Slightly disagree Neutral Slightly agree AgreeStrongly agree

While reading this chart, I found myself imagining things not directly related to what I was seeing in the chart.
 Strongly disagree Disagree Slightly disagree Neutral Slightly agree Agree
 Strongly agree

12. While reading this chart, I found myself generating new and original thoughts or ideas.

Strongly disagree Disagree Slightly disagree Neutral Slightly agree Agree Strongly agree

13. While reading this chart, I found myself concentrating on specific aspects or features of the chart.Strongly disagree Disagree Slightly disagree Neutral Slightly agree Agree

Strongly agree

14. While reading this chart, I had to pay attention to multiple things at the same time.Strongly disagree Disagree Slightly disagree Neutral Slightly agree AgreeStrongly agree

15. The content or message of this chart was interesting to me.Strongly disagree Disagree Slightly disagree Neutral Slightly agree AgreeStrongly agree

16. The features provided in this chart were interesting to me.Strongly disagree Disagree Slightly disagree Neutral Slightly agree AgreeStrongly agree

17. The look of this chart was novel and fresh.Strongly disagree Disagree Slightly disagree Neutral Slightly agree AgreeStrongly agree

18. While reading this chart, I experienced enjoyment from the chart in and of itself, and not because it was a means to an end.Strongly disagree Disagree Slightly disagree Neutral Slightly agree AgreeStrongly agree

19. I would want to read this chart if I saw it somewhere else and was not required to use it.Strongly disagree Disagree Slightly disagree Neutral Slightly agree AgreeStrongly agree

20. Overall, I am satisfied with this chart.Strongly disagree Disagree Slightly disagree Neutral Slightly agree AgreeStrongly agree

Next, I would like to learn about your opinion on the study.

1. To what extent is the information presented in a biased or unbiased manner?

Biased - 1 2 3 4 5 6 7 – Unbiased

2. How much you trust the study findings?Not at all -1 2 3 4 5 6 7-Completely

3. How credible is the study?Not at all -1 2 3 4 5 6 7-Completely

#### Scale Internal Consistency Analyses

**Chart Engagement (VisEngage)**: Based on Phase 1 responses, the 19 item Engagement scale had high internal consistency (Cronbach's alpha = 0.941, items = 19, N = 202).

#### B. Phase 3 Engagement Measure

I would like to learn about your opinion on how the study findings were presented. Overall, how much do you disagree or agree with the following statements?

While reading this presentation of findings, I found its look to be pleasing.
 Strongly disagree Disagree Slightly disagree Neutral Slightly agree Agree
 Strongly agree

The layout of this presentation of findings is clear and balanced.
 Strongly disagree Disagree Slightly disagree Neutral Slightly agree Agree
 Strongly agree

3. While reading this presentation of findings, I felt absorbed to the extent that I was not aware of my surroundings.

Strongly disagree Disagree Slightly disagree Neutral Slightly agree Agree Strongly agree

While reading this presentation of findings, time seemed to pass quickly.
 Strongly disagree Disagree Slightly disagree Neutral Slightly agree Agree
 Strongly agree

5. While reading this presentation of findings, I enjoyed and accepted any challenges it presented.

Strongly disagree Disagree Slightly disagree Neutral Slightly agree Agree Strongly agree

6. While reading this presentation of findings, I had to think carefully, deeply, or reflectively.

Strongly disagree Disagree Slightly disagree Neutral Slightly agree Agree Strongly agree

7. While reading this presentation of findings, I learned something that I had not known before (e.g., a new fact, concept, or piece of information).

Strongly disagree Disagree Slightly disagree Neutral Slightly agree Agree Strongly agree

I learned and figured out how to read this presentation of findingsalong the way.
 Strongly disagree Disagree Slightly disagree Neutral Slightly agree Agree
 Strongly agree

9. While reading this presentation of findings, I felt as though I was moving in or through it to learn about its content or message.

Strongly disagree Disagree Slightly disagree Neutral Slightly agree Agree

Strongly agree

10. While reading this presentation of findings, I was exploring its features and content in a gradual fashion.

Strongly disagree Disagree Slightly disagree Neutral Slightly agree Agree

Strongly agree

11. While reading this presentation of findings, I found myself imagining things not directly related to what I was seeing in the chart.

Strongly disagree Disagree Slightly disagree Neutral Slightly agree Agree

Strongly agree

12. While reading this presentation of findings, I found myself generating new and original thoughts or ideas.

Strongly disagree Disagree Slightly disagree Neutral Slightly agree Agree Strongly agree

13. While reading this presentation of findings, I found myself concentrating on specific aspects or features of the chart.

Strongly disagree Disagree Slightly disagree Neutral Slightly agree Agree Strongly agree

14. While reading this presentation of findings, I had to pay attention to multiple things at the same time.

Strongly disagree Disagree Slightly disagree Neutral Slightly agree Agree Strongly agree

15. The content or message of this presentation of findingswas interesting to me.Strongly disagree Disagree Slightly disagree Neutral Slightly agree AgreeStrongly agree

16. The features provided in this presentation of findingswere interesting to me.Strongly disagree Disagree Slightly disagree Neutral Slightly agree AgreeStrongly agree

17. The look of this presentation of findingswas novel and fresh.Strongly disagree Disagree Slightly disagree Neutral Slightly agree AgreeStrongly agree

18. While reading this presentation of findings, I experienced enjoyment from the chart in and of itself, and not because it was a means to an end.

Strongly disagree Disagree Slightly disagree Neutral Slightly agree Agree Strongly agree

19. I would want to read this presentation of findingsif I saw it somewhere else and was not required to use it.

Strongly disagree Disagree Slightly disagree Neutral Slightly agree Agree Strongly agree

1. Overall, I am satisfied with this presentation of findings.

Strongly disagree Disagree Slightly disagree Neutral Slightly agree Agree Strongly agree

2. To what extent is the information presented in a biased or unbiased manner?

Biased - 1 2 3 4 5 6 7 - Unbiased

Next, I would like to learn about your opinion on the study.

4. To what extent, if at all, are the study findings surprising?
Unsurprising - 1 2 3 4 5 6 7 - Surprising

5. How much do you trust the study findings?

Not at all -1 2 3 4 5 6 7-Completely

6. How credible is the study?

Not at all -1 2 3 4 5 6 7-Completely

**Chart Engagement (VisEngage)**: Based on Phase 3 responses of the overall sample, the 19 item Engagement scale had high internal consistency (Cronbach's alpha = 0.904, items = 19, N = 426).

## **Appendix I. Demographic Questionnaire**

Thank you for your time and input! In this last section, I would like to learn about you.

1. In a typical month, how often did you see data displayed in a chart?

Never Less than once a month Once a month Once a week Once a day More than once a day

2. In typical month, how often do you use data displayed in a chart to make decisions?

Never Less than once a month Once a month Once a week Once a day More than once a day

3. In a typical month, how often do you create a chart to display data?

Never Less than once a month Once a month Once a week Once a day More than once a day

- 4. What is the highest level of education you have completed?
  - A. Some High School
  - B. High School Diploma/GED
  - C. Associate's Degree/Some college
  - D. Bachelor's Degree
  - E. Master's Degree/Professional Degree
  - F. Doctorate/PhD
  - G. Other
  - H. Prefer not to answer
- 5. What is your age group?
  - A. 18 24 years old
  - B. 25 34 years old
  - C. 35-44 years old
  - D. 45 54 years old
  - E. 55-64 years old
  - F. 65 74 years old
  - G. Above 75 years old

- H. Prefer not to answer
- 6. What is your gender?
  - A. Male
  - B. Female
  - C. Other
  - D. Prefer not to answer

## **Appendix J. Behavioral Intent Survey**

Please mark the extent to which you would disagree or agree to engage in the following activities.

Tell a friend about the WISH program.
 Strongly disagree Disagree Slightly disagree Neutral Slightly agree Agree
 Strongly agree

Vote for a ballot measure to provide funding for the WISH program.
 Strongly disagree Disagree Slightly disagree Neutral Slightly agree Agree
 Strongly agree

3. Volunteer for the WISH program.

Strongly disagree Disagree Slightly disagree Neutral Slightly agree Agree Strongly agree

Recommend the WISH program to someone you know.
 Strongly disagree Disagree Slightly disagree Neutral Slightly agree Agree
 Strongly agree

## Scale Internal Consistency Analyses

Based on Phase 2 responses, the 4 item Behavioral Intention scale had high internal consistency (Cronbach's alpha = 0.874, items = 4, N = 340).
## Appendix K. Interactive Data Display

## A. Condition: Interactive Presentation with Strong Evidence

1. Prior to joining the WISH program, 10% of participants reported they were "extremely satisfied" with their work-life balance.

What percent do you think reported being "extremely satisfied" with their work life balance after completing the WISH program?



2. Your Guess: {Q1 Response Pipped Text}%

Research Finding: 55%

The percent of women who reported that they were "extremely satisfied" with their work-life balance increased from 10% to 55% among program participants.



-----page break------page break------

3. The researchers also collected data during the same timeframe from a comparison group of women who <u>did not</u> participate in the WISH program. At Time 1, 11% of the women in the comparison group reported feeling "extremely satisfied" with their work-life balance.

What percent of the women in the comparison group do you think reported being "extremely satisfied" with their work-life balance at Time 2?



4. Your Guess: {Q3 Response Pipped Text}%

Research Finding: 13%

The percent of women in the comparison group who reported that they were "extremely satisfied" with their work-life balance increased from 11% to 13%.



5. The researchers looked at whether participating in the WISH program had an impact on income.

What do you think was the median percent increase in income among women who participated in the WISH program?



6. Your Guess: {Q5 Response Pipped Text}%

Research Finding: 58%

Women who participated in the WISH program had a median 58% increase in their monthly income.



7. The researchers also collected data during the same timeframe from a comparison group of women who <u>did not</u> participate in the WISH program.

What do you think was the median percent increase in income among women who <u>did not</u> participate in the WISH program?

% 0-----100 -----page break------

8. Your Guess: {Q7 Response Pipped Text}%

Research Finding: 0%

Women in the comparison group did not have an increase in income during the study period.



## **B.** Condition: Interactive Presentation with Weak Evidence

1. Prior to joining the WISH program, 10% of participants reported they were "extremely satisfied" with their work-life balance.

What percent do you think reported being "extremely satisfied" with their work life balance after completing the WISH program?



2. Your Guess: {Q1 response pipped text}%

Research Finding: 15%

The percent of women who reported that they were "extremely satisfied" with their work-life balance increased from 10% to 15% among program participants.



3. The researchers also collected data during the same timeframe from a comparison group of women who <u>did not</u> participate in the WISH program. At Time 1, 11% of the women in the comparison group reported feeling "extremely satisfied" with their work-life balance.

What percent of the women in the comparison group do you think reported being "extremely satisfied" with their work-life balance at Time 2?

% 0-----100

4. Your Guess: {Q3 response pipped text}%

Research Finding: 13%

The percent of women in the comparison group who reported that they were "extremely satisfied" with their work-life balance increased from 11% to 13%.



-----page break------page break------

5. The researchers looked at whether participating in the WISH program had an impact on income.

What do you think was the median percent increase in income among women who participated in the WISH program?

% 0-----100

6. Your Guess: {Q5 response pipped text}%

Research Finding: 8%

Women who participated in the WISH program had a median 8% increase in their monthly income.



7. The researchers also collected data during the same timeframe from a comparison group of women who <u>did not participate</u> in the WISH program.

What do you think was the median percent increase in income among women who <u>did not</u> participate in the WISH program?

% 0-----100 ----page break-----

8. Your Guess: {Q7 response pipped text}%

Research Finding: 0%

Women in the comparison group did not have an increase in income during the study period.

