ERDC/ITL MP-21-8

Information Technology Laboratory

US Army Corps of Engineers®
Engineer Research and Development Center

ERDC
ENGINEER RESEARCH & DEVELOPMENT CENTER

Challenges in Evaluating Efficacy of Scientific Visualization for Usability and Aesthetics

Julie Baca, Daniel W. Carruth, Alex Calhoun, Michael Stephens, and Christopher Lewis

May 2021

Approved for public release; distribution is unlimited.

The U.S. Army Engineer Research and Development Center (ERDC) solves the nation's toughest engineering and environmental challenges. ERDC develops innovative solutions in civil and military engineering, geospatial sciences, water resources, and environmental sciences for the Army, the Department of Defense, civilian agencies, and our nation's public good. Find out more at www.erdc.usace.army.mil.

To search for other technical reports published by ERDC, visit the ERDC online library at https://erdclibrary.on.worldcat.org/discovery.

ERDC/ITL MP-21-8
May 2021

Challenges in Evaluating Efficacy of Scientific Visualization for Usability and Aesthetics

Julie Baca, Michael Stephens, and Christopher Lewis

Information Technology Laboratory
U.S. Army Engineer Research and Development Center
3909 Halls Ferry Road
Vicksburg, MS 39180

Daniel W. Carruth and Alex Calhoun

Center for Advanced Vehicular Systems
Mississippi State University
75 B. S. Hood Road
Starkville, MS 39762

Final report

Approved for public release; distribution is unlimited.

Prepared for U.S. Army Corps of Engineers
Washington, DC 20314-1000

Under BAA Contract Number W912HZ-17-C-0015, Task 4

Preface

This study was conducted for the U.S. Army Corps of Engineers (USACE), under BAA funding, Contract Number W912HZ-17-C-0015, Task 4.

The work was performed by the Supercomputing Research Center, U.S. Army Engineer Research and Development Center, Information Technology Laboratory (ERDC-ITL). At the time of publication of this Miscellaneous Paper, Mr. Robert Alter was DSRC Deputy Director and Mr. Robert Hunter was SRC Chief/DSRC Director. The Deputy Director of ERDC-ITL was Ms. Patti S. Duett and the Director was Dr. David A. Horner.

This paper was originally published online as part of the International Conference on Human-Computer Interaction on 4 July 2019 and published in *Design, User Experience, and Usability. Practice and Case Studies. Lecture Notes in Computer Science.*

The Commander of ERDC was COL Teresa A. Schlosser and the Director was Dr. David W. Pittman.

DISCLAIMER: The contents of this report are not to be used for advertising, publication, or promotional purposes. Citation of trade names does not constitute an official endorsement or approval of the use of such commercial products. All product names and trademarks cited are the property of their respective owners. The findings of this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

DESTROY THIS REPORT WHEN NO LONGER NEEDED. DO NOT RETURN IT TO THE ORIGINATOR.

Challenges in Evaluating Efficacy of Scientific Visualization for Usability and Aesthetics

Abstract. This paper presents the results of a study to evaluate the efficacy of scientific visualization for multiple categories of users, including both domain experts as well as users from the general public. Efficacy was evaluated for understanding, usability, and aesthetic value. Results indicate that aesthetics play a critical, but complex role in enhancing both user understanding and usability.

1 Introduction

Evaluating the efficacy of scientific visualization has long presented challenges to those working in the field, whether users, creators, or providers. Early research focused on honing specific visualization techniques and underlying algorithms [1,2]. As the technology matured, higher level user-centered efforts to evaluate scientific visualization began to appear [3-6].

A recent comprehensive review of evaluation practices in scientific visualization is given in [7]. The review encodes papers from 10 years of IEEE Visualization conferences to assess their evaluation practices. While reports of evaluation per se steadily rose during the review period, the same did not hold true with respect to user involvement in evaluations. Of the encoded papers, over 95% (955/1002) did not include users in a structured evaluation, but instead either 1) presented readers of papers with image or algorithm enhancements, asking them to assess any improvements for themselves or 2) relied on reports of feedback at informal demos to expert users [7].
Our study seeks to address these gaps by engaging multiple categories of potential actual users in formally evaluating the efficacy of scientific visualization.

2 Background

2.1 Evaluation Test Bed

Our work providing visualization services to scientists in the DoD High Performance Computing Modernization Program (HPCMP) motivated the need to evaluate scientific visualization from our users' many potential perspectives. We offer support to scientists analyzing large volumes of complex data in a variety of ways, including but not limited to assistance in using a visualization technique, as well as choosing which technique to use for a particular problem. Support may also involve a visualization specialist collaborating with the scientist in the use of the visualization program to extract data or images that highlight problems with an original computation.

Our users often need the visualization results to communicate and collaborate with other experts in their field to solve problems, but also to convey the results of their work to non-expert users who function as sponsors or liaisons to funding agencies. In addition, they must share their research for the purposes of outreach and education of the general public not trained in their research areas.

Hence, the two basic informational and communicative needs of our direct users can be broadly categorized as:

- Collaborating with other experts in their fields to conduct research;
- Communicating results to non-expert sponsors or public consumers

We designed our study to include both of these categories of users.

2.2 Related Research

Three types of evaluations conducted in the field of information visualization are relevant to our study and informed our approach:

1. Usability-centered
2. Aesthetic
3. Iterative, generative designed-based

Usability-centered evaluations assess ease-of-use through objective measurement of user task performance, as well as subjective measures of user satisfaction [8, 9]. Aesthetic evaluations focus specifically on measuring the user's perception of aesthetic quality of a graphic element or visualization, either through objectively quantifiable metrics, such as the number of bends or edges in a graph [10] or through personal judgments of beauty [11].

Iterative, designed-based evaluations use any results gained in an iterative feedback loop to refine or enhance a visualization or other UI as it is being developed [12,13]. Any of these categories of evaluations may overlap or be used in parallel in a given study, as for example in [14,15] to examine effects of aesthetics on usability. Of particular relevance to our research, the aesthetic evaluations in these studies based certain key hypotheses of expected user responses on the work of Tufte [16] who argued that the human eye finds nature's palette most harmonious, and that blues, greens, and browns would have the most desirable impact in information display. Additionally, the presence of any organic, lifelike movement or animation, such as that found in nature, would provide further benefit.

Our study uses aspects of evaluation types 1 and 2, with the goal of using the results going forward in the type 3 designed-based evaluations that allow enhancement of a visualization as it is developing.

3 Methods

Approximately 30 users from each category, expert and non-expert, participated in the study. Participants were recruited from among faculty, general staff and students at a university research center. They were asked to watch videos of a scientific visualization, answer questions about its content, and to evaluate its aesthetic quality. They were also asked to rate their perception of the impact of the work visualized for science and the military.

3.1 Visualization Description

The Army is studying heavy fuel engines relying exclusively on direct injection fuel delivery systems. To meet requirements, the engines must significantly advance current fuel flexibility and fuel conversion efficiencies. In propulsion systems, energy conversion by combustion relies on the use of jet fuel in the compressed liquid form. The initial step in the energy conversion process is the atomization, or disintegration of the coherent liquid core, which significantly impacts the droplet-size distribution and fuel conversion efficiency. In combusting scenarios, the liquid fuel must be fully atomized, evaporated and mixed with the carrier gas-phase environment. Hence, the interaction of liquid atomization and spray vaporization is critical, as it determines the level of energy and fuel supplied to the flame.

Also relevant to our work were emails exchanged with the visualization staff regarding the principal investigator's intent for his audience in creating the visualization, as shown in the excerpt below:

> "In the visualization, I'd like to highlight the atomization breakup features of the spray, that is, surface instability corrugations, ligament formation, spray development. Perhaps a transparent media (color) would also show the internal flow structure. I'm open to your expert suggestions as well."

Users in the study were shown an animation of the visualization of the atomization spray that was subsequently created iteratively with our visualization team and the researcher. Key frames of the animation illustrating the atomization process are shown in Figures 1-5.

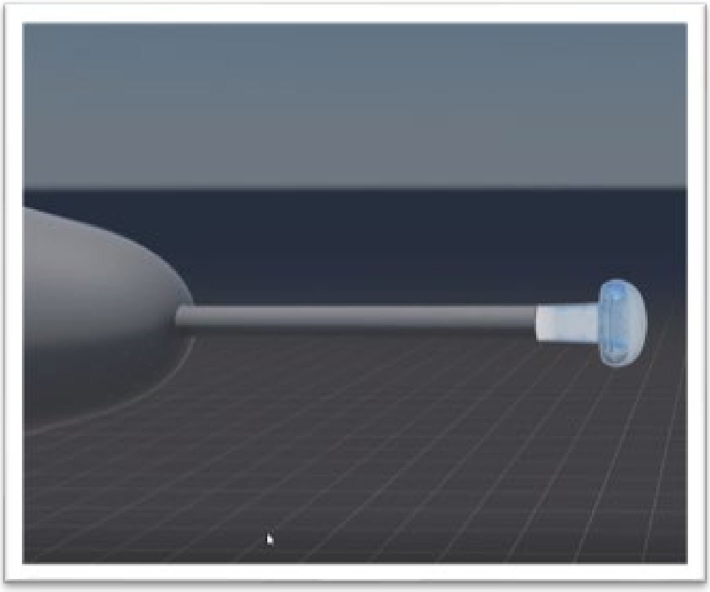

Fig. 1. Early frame in energy conversion process

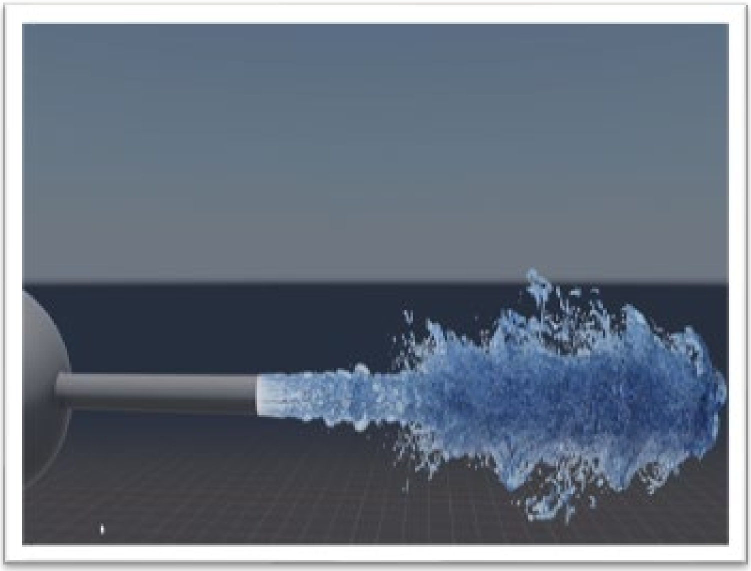

Fig. 2. Atomization spray droplets beginning to form

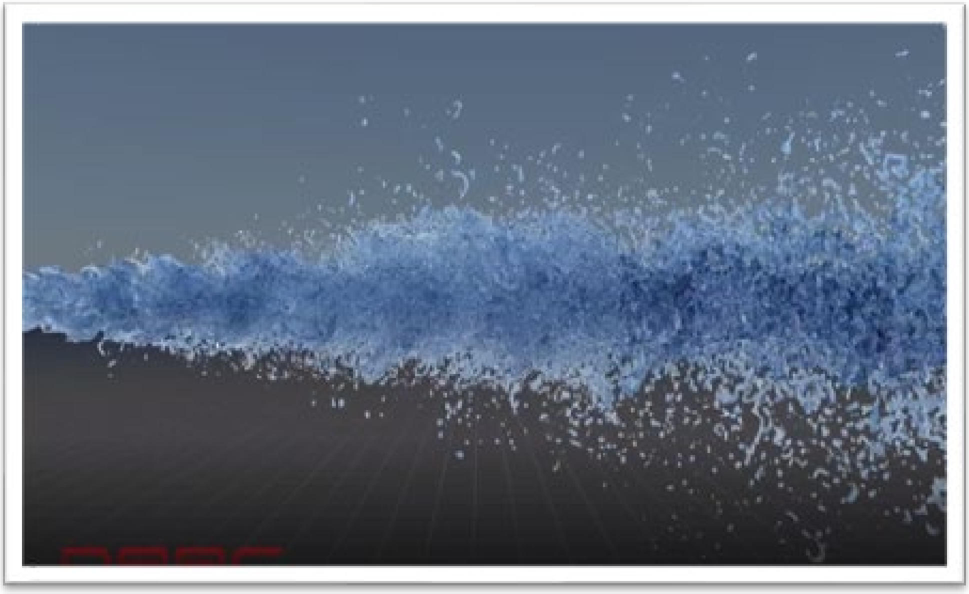

Fig. 3. Atomization spray mid-animation

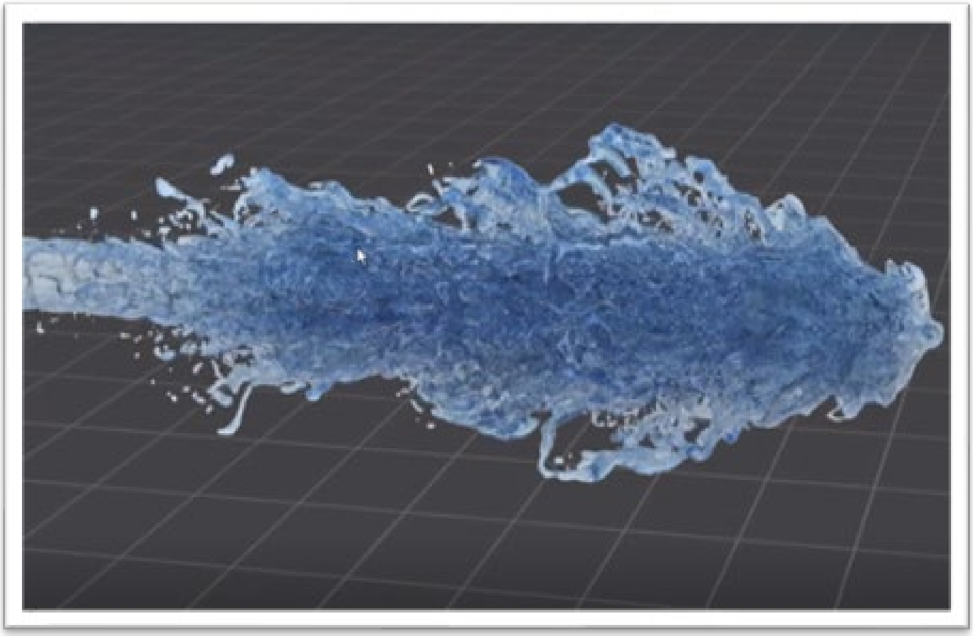

Fig. 4. Atomization spray later in animation

Fig. 5. Atomization spray droplets final animation frame

3.2 Experimental Treatments

Users viewed the set of visualizations of the atomization and spray vaporization interactions shown in Figures 1-5. They were also given a textual description of the entire process on the first screen shown. There were no "tasks" for users to perform in the traditional usability sense; however they were asked to answer questions about what they viewed regarding content, impact and aesthetics. They were also asked to rate their confidence in their answers

Once they advanced beyond the first screen to answer questions, users were allowed to view the visualizations as many times as they wished, but could no longer view the textual description for the remainder of the experiment in order to isolate the effectiveness of the visualization itself. The relative completeness of their answers, and their confidence ratings, functioned as measures of task completion. The survey contained 8 questions regarding scientific content, its potential importance to science and the military, with 3 questions regarding users' confidence in their answers.

Users were asked in 2 additional questions to rate their perceptions of aesthetic quality of the visualizations and to give an explanation of their rating. Based on Tufte's assumptions [16], tested in [15], we expected that the spectrum of blues used in the visualization of the atomized spray, its similarity to the appearance of water, and the organic lifelike movement of the animation of the spray, would have a positive effect on user ratings and explanations.

4 Results

4.1 Scoring Answers

Scoring understanding of textual answers (i.e., non-numeric) involves a subjective element. Multiple readers (3) provided scores for each answer. Inter-rater reliability for the 3 readers was calculated using Gwet's AC2 [17] and is reported in Table 1. A final score was calculated by taking the average of the 3 scores.

Answers were scored on the following 3-point scale:

0 Wrong, absent, or irrelevant
1 Correct undetailed
2 Correct detailed

Overall results for accuracy of textual answers for the central problem, the main idea, and the impact of the research, as well as the user's confidence in their own answers and ratings is given in Table 1.

Table 1. Overall results for accuracy and user confidence.

	Accuracy (0-2)	Confidence (1-5)	Gwet AC2
Central Problem	0.89	2.55	.49
Main Idea	1.08	2.78	.63
Impact	0.85	2.48	.39

Overall scores for other questions with numeric ratings that did not require users to write a descriptive answer included significance to science, importance to military, and perceived aesthetics. These scores are shown in Table 2-3.

Table 2. Overall results for significance to science and importance to military

	Least to Highest (1-5)
Significance to Science	3.63
Importance to Military	4.3

Table 3. Overall results for perceived aesthetics

	Ugly to Beautiful (0-100)
Perceived aesthetics	72.7

4.2 Analysis

Several key questions motivated our research and guided our analysis of the results. Discussion and analysis for each question are presented below.

1. Does perceived aesthetics predict perception of the significance of the research to science?
Perceived aesthetics was correlated with individuals' assessment of the significance of the research to science, rho = .432, $p < .01$. Participants who perceived the animation as more attractive were more likely to perceive the research as significant to science. An analysis of variance indicated that for every point towards beautiful, there was a .37 increase in significance, $F(1,39) = 6.05$, $p < .05$.

2. Does perceived aesthetics affect understanding?
Participant understanding of the impact of the research was also correlated with their perception of the aesthetics of the video, rho = -.356, $p < .05$. Participants that felt the visualization was less attractive were more likely to provide an accurate assessment of impact.

3. Does perception of aesthetics affect confidence in responses?
The analysis revealed no evidence of a relationship between the aesthetics of the animation and participant self-reported confidence in their responses.

4. Does self-reported prior knowledge of underlying science or modeling and simulation predict accuracy of answers?
Participant self-reported prior knowledge did not predict understanding of the central problem ($p = .34$), impact ($p = .31$), or the main ideas presented ($p = .78$).

5. Does self-reported prior knowledge of underlying science or modeling and simulation predict confidence?

An analysis of variance indicates that self-reported knowledge of modeling and simulation combined with their understanding of the central problem does predict self-reported confidence in their understanding of the central problem $F(4,39) = 2.92, p < .05$). By itself, there was no relationship between self-reported knowledge and the accuracy of their responses. There was no evidence that this relationship also applied to confidence in their understanding of the impact of the research, $F(4,39) = 1.13, p = .36$), or of the main ideas presented in the video, $F(4,39) = 1.98, p = .12$.

6. Is understanding related to confidence in responses?
Participant confidence in their understanding of the central problem was correlated with accuracy, rho = -.342, $p < .05$, but participants with higher confidence in responses were actually less accurate. However, the analysis did not provide support for a relationship between confidence in understanding of the main ideas and the impact and accuracy in responses.

7. Does understanding predict perceived importance to the military?
Understanding of the central problem was negatively associated ($B = -.64$) with the perceived importance to the military (better understanding led to less perceived importance), $F(3,39) = 3.29, p < .05$.

In contrast, understanding of the main ideas presented in the visualization was positively associated ($B = .48$) with perceived importance of the research to military.

8. Does understanding predict perceived significance to science?
The accuracy of participants' understanding of the central problem, main ideas, and impact were not significant predictors of how significant the participants perceived the research to be to science.

9. Will adherence to Tufte's guidelines on the use of organic qualities and colors found in nature positively affect user perceived aesthetics?
Users' explanations for their aesthetic ratings contained a relatively high number of positive comments (18/41, roughly 43%) related to the color choice (blue) and/or the realistic, lifelike quality of the animation of the liquid, including its similarity to water. Given the relatively high overall aesthetic rating of (72.7 points out of 100) this appears to support Tufte's recommendations. Examples of user comments are presented below:

> U1: "…easy on the eyes to visually understand where the fuel is going. It has good colors."
>
> U2: "The graphical imagery was realistic and the colors popped. The last shot of the up-close visualization was stunning."

U3: "I liked how the fuel was shot in slow motion and the many different views of that process given. …The coloring was also nice."

U4: "…the blues and 3D patterns were cool."

U5: "…the colors and detail of the liquid are attractive.

U6: "…the blue colors of the fuel were fairly visually pleasing and seemed to demonstrate the atomization and injection."

U7: "…the color choices help to distinguish the different regions of density process well."

5 Conclusions

The expectation that aesthetics might play a critical role in both user understanding of the content of the visualization as well as its usability was confirmed, but in more complex patterns than expected. Perceived aesthetics proved a predictor of user perception of the significance of research to science, i.e., "uglier" aesthetic ratings were correlated with lower user perceptions of research significance. In addition, perceived aesthetics affected user understanding of the impact of the research; however, higher or "more beautiful" aesthetic ratings were correlated with less accurate assessments of impact. However, aesthetics were not a significant predictor of user confidence in their ratings. Finally, tracking with Tufte's assertions [16], users' explanations for higher aesthetic ratings included a relatively high number of comments (18/41, roughly 43%) related to the choice of natural, lifelike movement, i.e., water, and/or the colors from nature, i.e., the color blue.

Other key questions of the research regarding self-reported prior knowledge proved complex as well. While this factor did not predict user understanding, it was a predictor of user confidence, i.e. those with self-reported prior knowledge or expertise were more likely to report a high level of confidence in their ratings.

We plan to continue the research by expanding the study in several ways. First we plan to conduct another study using visualizations providing more contextual information such as conceptual animations, and musical tracks to assess the level at which these factors may affect the key questions regarding usability and understanding. Second, we plan to include comparison studies of the original researcher's visualization, which did not adhere to Tuft's guidelines, against the visualization created in collaboration with our visualization team.

Our long-term goal is to use the results of these studies during design and development to enhance the quality of visualizations provided to researchers, scientists, and the general public. This research will enable a more explicit formulation of a visualization usability process to follow to attain our goal.

Acknowledgements

This material is based upon work supported by the U.S. Army TACOM Life Cycle Command under Contract No. W56HZV-08-C-0236.

References

1. Wong, C. and Bergeron, D.: 30 years of multidimensional multivariate visualization. [book auth.] G. Hagen, H. and Muller, H. Nielsen. *Scientific Visualization.* Los Alamitos,CA, USA, IEEE Computer Society (1997).
2. Rheuter, L., Tukey, P., Maloney, L.T., Pani, J.R., Smith, S.: Human perception and visualization. Viz'90 Proceedings of the 1st Conference on Visualization '90, pp. 401-406, IEEE Computer Society Press, San Francisco, CA, USA (1990).
3. Herndon, K., van Dam, A., and Gleicher, M: The challenges of 3D interaction: A CHI '94 Workshop. SIGCHI Bulletin 26 (4), 1-9 (2007).
4. Komlodi, A. Rehingans, P., Ayachit, U., Goodall, J., and Joshi, A.: A user-centered look at glyph-based security visualization. IEEE 2005 Workshop on Visualization for Computer Security, pp. 21-28. IEEE, Minneapolis, MN, USA (2005).
5. Santos, B.S., Silva, S., Teixeira, L., Ferreira, C., Dias, P., Madeira, J.: Preliminary usability evaluation of PolyMeCo: A visualization based tool for mesh analysis and comparison. Geometric Modeling and Imaging '07, pp. 133-139. IEEE, Zurich, Switzerland IEEE (2007).
6. Tory, M., Moller, T.: Evaluating visualizations: Do expert reviews work? IEEE Visualization Viewpoint in Computer Graphics and Applications. 25, 8-11 (2005).
7. Isenberg, T., Isenberg, P., Chen, J., Sedlmair, M., and Moller, T. A systematic review on the practice of evaluating visualization, IEEE Transactions on Visualization and Computer Graphics. 19(12), 2818-2827 (2013).
8. Nielsen, J. Usability Inspection Methods. New York: John Wiley & Sons, (1994).
9. Norman, D.A., and S.W. Draper. User Centered System Design: New Perspectives on Human Computer Interaction. Hillsdale, NJ: Erlbaum (1986).
10. H. Purchase, J.A. Allder, and D. Carrington: Metrics for graph drawing aesthetics. .Journal of Visual Languages and Computing 13, 501-516, (2002).
11. Hartmann, J.: Assessing the attractiveness of interactive systems. In CHI '06 Extended Abstracts on Human Factors in Computing Systems, pp. 1755-1758. ACM Press, Montreal, Quebec, Canada (2006).
12. Preece, J., Rogers, Y., and Sharp, H.: The process of interaction design. Chap. In Interaction Design: Beyond Human-Computer Interaction. John Wiley & Sons, New York, NY (2002).
13. Jackson, B., Coffeey, D., Thorsen, L., Schroeder, D., Ellingson, A.M., Nuckley, D.J., and D.F. Keefe.: Towards mixed method evaluations of scientific visualizations and design process as an evaluation tool. BELIV 2012, Seattle Washington, USA (2012).

14. Cawthon, N. and A. Vande Moere: Qualities of perceived aesthetic in data visualization. CHI 2007, pp. 1-11. ACM Press, San Jose, USA (2007).
15. Cawthon, N. and A. Vande Moere: The effect of aesthetic on the usability of data visualization. 11th International Conference Information Visualization IV'07, pp. 637-648, IEEE Computer Society, Washington D.C., USA (2007).
16. E. Tuft:, Envisioning Information. Graphics Press, Cheshire, Conn. U.S.A. (1990).
17. K.L. Gwet: Handbook of Inter-Rater Reliability (3rd Ed.). Advanced Analytics, LLC, Maryland. U.S.A. (2012).

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

1. REPORT DATE *(DD-MM-YYYY)* May 2021	**2. REPORT TYPE** Final	**3. DATES COVERED** *(From - To)*
4. TITLE AND SUBTITLE Challenges in Evaluating Efficacy of Scientific Visualization for Usability and Aesthetics		**5a. CONTRACT NUMBER** W912HZ-17-C-0015 **5b. GRANT NUMBER** **5c. PROGRAM ELEMENT NUMBER**
6. AUTHOR(S) Julie Baca, Daniel W. Carruth, Alex Calhoun, Michael Stephens, and Christopher Lewis		**5d. PROJECT NUMBER** **5e. TASK NUMBER** 4 **5f. WORK UNIT NUMBER**
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) U.S. Army Engineer Research & Development Ctr Information Technology Laboratory 3909 Halls Ferry Road Vicksburg, MS 39180	Center for Advanced Vehicular Systems Mississippi State University Starkville, MS 39762	**8. PERFORMING ORGANIZATION REPORT NUMBER** ERDC/ITL MP-21-8
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) Headquarters, U.S. Army Corps of Engineers 441 G Street NW Washington, DC 20314-1000		**10. SPONSOR/MONITOR'S ACRONYM(S)** **11. SPONSOR/MONITOR'S REPORT NUMBER(S)**

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

This paper was originally published online as part of the *International Conference on Human-Computer Interaction* on 4 July 2019.

14. ABSTRACT

This paper presents the results of a study to evaluate the efficacy of scientific visualization for multiple categories of users, including both domain experts as well as users from the general public. Efficacy was evaluated for understanding, usability, and aesthetic value. Results indicate that aesthetics play a critical, but complex role in enhancing both user understanding and usability.

15. SUBJECT TERMS

Scientific visualization, Usability, Aesthetics

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			**19b. TELEPHONE NUMBER** *(include area code)*
Unclassified	Unclassified	Unclassified	SAR	15	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. 239.18