J Vis (2025) 28:645–659
https://doi.org/10.1007/s12650-025-01051-y

REGULAR PAPER

**Nina Errey · Didar Zowghi · Tuck Wah Leong · Yadong Wu · Xiaoru Yuan · Weidong Huang · Christy Jie Liang**

# Heuristics for evaluating narrative visualization: a validation study

Received: 15 October 2024 / Revised: 6 January 2025 / Accepted: 20 January 2025 / Published online: 24 February 2025


**Abstract** Narrative visualization integrates data visualization and narrative techniques to convey a compelling story. Narrative visualization is notoriously difficult to evaluate. One solution is heuristic evaluation, using a domain-specific set of heuristics. This paper validates a set of heuristics proposed specifically for evaluating narrative visualization. We conducted studies with experienced narrative visualization practitioners in both summative and formative settings. We found that the set of heuristics showed promise in a summative setting, where similar responses evidenced that the set of heuristics could provide reliable evaluation metrics. Furthermore, in a formative setting, implementing the set of heuristics was reported to be useful in the design process; however, due to their limited focus, we recommend that it be implemented in conjunction with other evaluation guidelines.

Didar Zowghi, Tuck Wah Leong, Yadong Wu, Xiaoru Yuan, Weidong Huang and Christy Jie Liang have contributed equally to this work.

**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1007/s12650-025-01051-y.

N. Errey (✉) · C. J. Liang
School of Computer Science, UTS, Sydney, Australia
E-mail: nina.errey@uts.edu.au

C. J. Liang
E-mail: jie.liang@uts.edu.au

D. Zowghi
Data61, CSIRO, Sydney, Australia
E-mail: didar.zowghi@data61.csiro.au

T. W. Leong
School of Computing Technologies, RMIT, Melbourne, Australia
E-mail: tuck.wah.leong@rmit.edu.au

Y. Wu
Computer Science and Engineering, Sichuan University of Science and Engineering, Yibin, China
E-mail: wyd028@163.com

X. Yuan
School of Intelligence Science and Technology, Peking University, Beijing, China
E-mail: xiaoru.yuan@pku.edu.cn

W. Huang
Transdisciplinary School, UTS, Sydney, Australia
E-mail: weidong.huang@uts.edu.au

**Keywords** Human-centered computing · Information visualization · Visualization · Visualization application domains · Empirical studies in visualization

## 1 Introduction

The capacity for narrative visualization to convey complex, important information compellingly has resulted in it being an increasingly popular communication medium. It has been employed in various contexts, ranging from explaining the extinction crisis to America's debt ceiling (Janicki et al. 2022; Lee et al. 2023). However, the means by which practitioners can conduct rigorous evaluations of their narrative visualization remain problematic.

Conventional empirical evaluation methods and metrics are insufficient in evaluating narrative visualization's experiential nature. Narrative visualization practitioners use narrative techniques, appropriated from classical narrative disciplines such as literature, comics, and film (Hullman and Diakopoulos 2011; Amini et al. 2015; Bach et al. 2017). Such techniques can create rich, affective reader experiences that help create a story in the reader's mind (Ryan 2004). Empirical studies are needed in narrative visualization evaluation as it is a distinctive form of visualization that bridges the divide between traditional storytelling and the disparate technicalities of data visualization.

One possible method to evaluate narrative visualization is heuristic evaluation. First described by Nielsen in 1994, it involves usability experts who review a system and judge how well it meets the objectives of a list of predefined guidelines—heuristics (Nielsen 1994). The visualization community has recognized that heuristic evaluation could be a valuable tool for evaluation (Tarrell et al. 2014; Carpendale 2008). Indeed, various sets of heuristics have been proposed for information visualization (Forsell and Johanson 2010; Zuk et al. 2006; Santos et al. 2015). Recently, a set of heuristics specifically for evaluating narrative visualization was proposed (Errey et al. 2023). Unlike other visualization heuristic sets, this set incorporates the storytelling aspects of narrative visualization. It was developed from a survey of practitioners, who were integral in their establishment. This heuristic set has yet to be assessed as useful to practitioners when evaluating narrative visualization.

In this paper, we describe how we validated a narrative visualization specific set of heuristics and outline the results of this validation. We conducted studies in both summative and formative settings to ascertain whether the heuristics could be useful in different contexts and at different stages in the development cycle. Summative evaluation is described as a 'seal of approval' when evaluating a visualization (Ellis and Dix 2006). In our summative study, we tasked multiple evaluators to individually evaluate a published or a working prototype of a narrative visualization (Nielsen 1994). The formative study differs in the development phase. Occurring much earlier in the project cycle, our formative study aimed to investigate whether the set of heuristics can assist with the early design of the narrative visualization. In this study, we tasked practitioners to create a wireframe for narrative visualization. Wireframe creation is a common step in the formative stages of data visualization development. We found that when implemented in a summative setting, the set of heuristics resulted in a good level of inter-rater reliability in evaluator responses. Consistent evaluation responses meant that this evaluation method could potentially standardize narrative visualization evaluation as it evidences a reliable evaluation method. In a formative setting, practitioners reported the heuristic set useful. This was because it provided structure to their design process. In both studies, we found that practitioners mentioned that they would integrate other guidelines or heuristics to complement the heuristic set. They reasoned that the domain-specific set of heuristics was too narrowly focused on the storytelling aspects of narrative visualization. With this research, we contribute an important step toward empirically validating a set of heuristics for narrative visualization evaluation. To conclude this paper, we identify opportunities for further refinement of the heuristics and future research related to narrative visualization evaluation.

## 2 Related work

In this section, we define narrative visualization, summarize recent research, and explore the challenges in its evaluation. We then describe domain-specific heuristic evaluation and its relationship with visualization. Finally, we motivate our work by describing the importance of validating domain-specific heuristics.

### 2.1 Narrative visualization

Narrative visualization is defined as a story consisting primarily of visualization steps, which can include text and images but essentially is based on data (Kosara and Mackinlay 2013). Segel and Heer were the first to coin the term and map the narrative visualization design space (Segel and Heer 2010). They categorized narrative visualization into overarching genres, such as comic strips or slide shows. Narrative visualization research has further detailed aspects of the design space, including story sequences, rhetorical strategies and factors that encourage reader engagement (Hullman, et al. 2013; Hullman and Diakopoulos 2011; McKenna et al. 2017). While delineating the unique and complex characteristics of narrative visualization, more understanding is required into how it can be practically evaluated.

In an attempt to capture the richness and complexity of narrative visualization, novel evaluation methods have been investigated. Some examples of novel evaluation methods studied include elicitation interviews, focus groups, and reaction cards (Nowak et al. 2018; Figueiras 2014; Merčun 2014). The aforementioned studies all incorporate end-users. Many narrative visualization practitioners do not employ end-user testing in their development process (Errey et al. 2023). This is largely due to end-user testing related drawbacks such as being expensive, laborious, and time-consuming. One evaluation method that does not include end-users is heuristic evaluation. This benefit, among others, gives us reason to believe that heuristic evaluation could be a promising solution to the challenge of practically evaluating narrative visualization.

### 2.2 Heuristic evaluation for visualization

Initially introduced by Nielsen, heuristic evaluation is an inspection method of evaluation (Nielsen 1994). Inspection methods of evaluation are a group of evaluation methods, where a small number of experts inspect a user interface to discover gaps in its design. Heuristic evaluation differs from other inspection methods because it incorporates a list of guidelines or heuristics in the evaluation procedure. Nielsen described heuristic evaluation as a 'discount' evaluation method as it is quick to perform and does not require end-users (Nielsen 1994). This method is said to be particularly helpful early in the design process, however, it can be used at all stages of development. Heuristic evaluation is most effective when the heuristics are domain-specific. Similar to other domains, domain-specific heuristics are recommended for visualization evaluation (Forsell and Johanson 2010; Zuk et al. 2006).

Tory and Möller were the first to empirically establish that expert reviews, using a set of heuristics or guidelines, are a valuable tool in the domain of visualization (Tory and Möller 2005). This is because heuristic evaluation can bridge the gap between visualization design and technical implementation (Kosara 2007; Brath and Banissi 2016). Another benefit is that heuristic evaluation can provide structure to the visualization evaluation process (Tory and Möller 2005). Furthermore, methodologies that adopt a heuristic approach are evidenced to be useful for visualization evaluation. For example, the colloquially named 'ICE-T' methodology was found to be promising when valuing a visualization (Wall et al. 2019). Alternatively, combining heuristics with other methods such as question-based scoring, has been found to be a robust visualization evaluation approach (Hearst et al. 2016). Another usage for heuristics is as an effective aid to educate students about the concepts and foundations of visualization (Santos et al. 2016).

Finding an appropriate set of heuristics is the main challenge for visualization evaluation (Zuk et al. 2006). Multiple sets of heuristics exist for visualization evaluation, each with a different intent and focus. For example, Amar and Stasko identify heuristics designed to cover known gaps in visual analytics processes (Amar and Stasko 2004). To assess a visualization's usability, Forsell and Johansson compiled a list of published heuristics and tested them on a collection of usability problems to identify 10 heuristics for information visualization usability evaluation (Forsell and Johanson 2010). All aforementioned heuristics are not specific to narrative visualization evaluation. While evaluation criteria and design guidelines have been proposed for narrative visualization they are not designed primarily for heuristic evaluation (Amini et al. 2015; Nussbaumer Knaflic 2015). It is therefore we focus on a set of heuristics that are specifically proposed for narrative visualization evaluation (Errey et al. 2023).

Carpendale described multiple disadvantages of heuristic evaluation for visualization. For example, the heuristics may be neither easily understandable nor straightforwardly applicable by practitioners who are not experienced evaluators. (Carpendale 2008). Moreover, being bounded by a set of rules could result in heuristics stifling the creativity of the designer (Carpendale 2008). Another notable issue is that domain-specific heuristics are narrowly focused meaning usability problems may be overlooked (Carpendale 2008).

Such limitations in heuristic evaluation led Nielsen to recommend that heuristic evaluation should, optimally, complement end-user testing (Nielsen 1994).

One barrier to adopting heuristics is translating the heuristic evaluation method into a realistic operational process (Jeffries et al. 1991). Practical difficulties arise, such as time constraints (Jeffries et al. 1991). In the early phases of development, clear objectives and variables might not yet be well defined. Effective heuristics function as a way of focusing attention on important aspects of the visualization and saving time by adding structure to evaluation (Tory and Möller 2005). In this work, we adopt Scriven's definition of formative evaluation (Scriven 1967). Fundamentally, formative evaluation means detecting deficiencies in an early to intermediate version of an application to improve its design. Carpendale described how heuristics can act as a checklist to remember during the formative stages of visualization development (Carpendale 2008). In contrast to formative evaluation, summative evaluation aims to assess or compare a post-deployment or a near deployment prototype (Scriven 1967). Summative evaluation is necessitated in contexts such as industry awards or student marking. The results of the summative evaluation must be reliable so that they can be objective and repeatable (Andrews 2006).

### 2.3 Validating domain-specific heuristics

Validating domain-specific heuristics is crucial because, without empirical validation, weaknesses in the heuristics might not be identified. Weak heuristics can result in improper or poorly evaluated applications. Hermati et al. systematically reviewed domain-specific heuristics (Hermawati and Lawson 2016). They found a documented lack of empirical validation in establishing a set of domain-specific heuristics.

Formalized methods for the validation of heuristics are not yet widely implemented (Hermawati and Lawson 2016). Hermati et al. found that the most common form of validation consisted of applying the heuristics by experts and analyzing their outcomes. Other validation methods included comparing the outcomes of heuristic evaluation to end-user evaluation outcomes. In this study, we have adopted the former method of validation. A similar method to validating heuristics was used in multiple visualization evaluation studies (Wall et al. 2019; Hearst et al. 2016). Therefore, our validation approach consisted of applying the heuristics by experts and analyzing their outcomes. We focus on the usefulness and practicality of the heuristic set. Using real- world examples, settings and practitioners we are able to bridge the divide between academia and visualization practice.

## 3 Research method

This work aimed to validate a set of heuristics for evaluating narrative visualization. We performed two studies in two varied evaluation settings to achieve this aim. To see raw data, please visit the supplementary material: https://osf.io/6aytu/?view_only=33469c21cfdd40b68b14dd910c7af6a0.

### 3.1 The set of heuristics

We directly borrowed the set of heuristics from the original work, where we made no modification to the heuristic framework structure or the individual heuristic intent (Errey et al. 2023). We adapted the set of heuristics into a series of questions, which is common approach in heuristic evaluation. Similar adaptations were successful in evaluating programming languages (Sadowski and Kurniawan 2011). Please see Table 1 for the set of heuristics adapted into questions.

The development of the set of heuristics was a three phase process (Errey et al. 2023). These phases, in order of progression were; literature review, practitioner survey and finally practitioner interviews. The findings from the literature review were synthesized into a set of foundational topics featured in a survey of practitioners. The survey results then influenced interview questions, where the set of heuristics formed the basis of discussion. This iterative process of developing a set of heuristics is described in visualization evaluation literature (Carpendale 2008).

The heuristic set was minimal in its design, using a minimalistic approach similar to Nielsen's original usability heuristic set (Nielsen 1994). The set is split into three overarching categories; 'composition', 'reader experience' and 'credibility and trust.' The 'composition' category relates to the visual aesthetic, information distribution, and overall layout. The 'reader experience' category relates to how a reader

**Table 1** The set of heuristics for evaluating narrative visualization adapted into questions

| QID | Category | Question | Vis A | | Vis B | | Vis C | |
|---|---|---|---|---|---|---|---|---|
| | | | M | SD | M | SD | M | SD |
| Q1 | Composition | Does the narrative visualization flow logically? | 3.58 | 1.08 | 4 | 0.9 | 2.94 | 1.1 |
| Q2 | | Does the narrative visualization slowly guide the reader into complexity? | 3.25 | 1.35 | 3.17 | 1.04 | 2.35 | 1.02 |
| Q3 | | Do the colors of the narrative visualization have any unwanted cultural or emotional connotations? | 2.17 | 1.09 | 2.17 | 1.24 | 2.23 | 1 |
| Q4 | | Does the text, specifically the title, aid the reader in understanding the narrative visualization and intrigue the reader to read further? | 3.88 | 0.89 | 4.23 | 0.87 | 2.7 | 1.36 |
| Q5 | | Does all data visualized relate to the focus of the narrative visualization? | 4.11 | 0.75 | 4.52 | 0.49 | 3.76 | 1.05 |
| Q6 | | Does the narrative visualization retain interest without being too long? | 3.88 | 0.67 | 3.7 | 1.22 | 3.23 | 1.39 |
| Q7 | Reader Experience | Does the narrative visualization function without the need for interactivity? | 3.05 | 1.21 | 3.47 | 1.53 | 4.17 | 0.78 |
| Q8 | | Does the narrative visualization personally relate to the reader's frame of reference? | 2.88 | 1.13 | 3.82 | 0.78 | 3.41 | 1.08 |
| Q9 | | Does the narrative visualization use recognizable universal metaphors? | 2.58 | 0.97 | 3.88 | 0.96 | 2.76 | 1.16 |
| Q10 | Credibility and Trust | Does the narrative visualization identify a data source? | 4.29 | 1.17 | 4.35 | 1.13 | 1.52 | 0.97 |
| Q11 | | Does the narrative visualization present the data in a confusing manner that could lead to misinterpretation from the reader? | 3.05 | 0.93 | 2.41 | 1.23 | 3.41 | 0.77 |

The first column contains the question ID, the second column contains the upper-level category, the third column contains the questions, and the last three columns represent each visualization example studied in the 'summative study' with correlated mean and standard deviation

interacts with and experiences narrative visualization. Finally, 'credibility and trust' relates to data quality and ability for the narrative visualization to be believed.

## 3.2 Summative setting study research method

The aim of this first study was to ascertain if a set of heuristics could be used to evaluate narrative visualization with reliable outcomes when evaluating a post-deployment or a near deployment prototype. A similar approach has been used for visualization research in the past (Wall et al. 2019; Hearst et al. 2016). We also asked practitioners about the usefulness and feedback on the set of heuristics.

### *3.2.1 Procedure*

This experiment consisted of three phases. The introduction phase, the evaluation phase, and lastly the qualitative feedback phase. See Fig. 1 for a diagram of the experiment procedure. Our survey instrument was hosted on the Qualtrics survey platform (Qualtrics 2024). During the introduction phase, the participant was first presented with a consent form, where ethics information was provided. If the participants consented, they would be asked demographic questions, including gender, years worked, and designation. After completing the demographic questions, the participant was presented with a set of heuristics adapted into questions, with a short explanation attached to each heuristic. Please see Table 1 for the questions presented in the introductory phase, which were also used during the evaluation phase of the experiment.

The second phase, the evaluation phase, presented the participants with three separate narrative visualization examples. We used a pseudorandom order of visualizations to minimize potential ordering effects. The participant could leave feedback after applying each heuristic. Alternatively, they could provide overall feedback at the end of the experiment. The feedback questions were not mandatory to answer. All other questions were mandatory.

### *3.2.2 Materials*

Three narrative visualizations were selected as examples in our study. The rationale for selecting each example was primarily based on if we received written permission from their authors to allow us to use their work. We selected two exemplary examples and one that the authors of this paper developed. The third

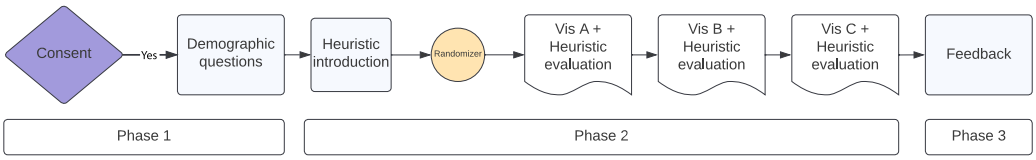


**Fig. 1** Procedure for summative study experiment. Phase 1 consisted of consent and demographic questions. Phase 2 consisted of a heuristic introduction, random order of three visualizations followed by heuristic evaluation. Phase 3 consisted of qualitative feedback questions

example acted as an ineffective example. We name each Visualization A, Visualization B, and Visualization C.

Visualization A won the Information is Beautiful Challenge award for best inter- active visualization (Information is Beautiful Awards 2014). Ti tled "Has the cost of a human life changed?" by Carlo Zapponi, it compared the cost of human life over the centuries (Zapponi 2014). Visualization A could be described as a 'martini glass' structured narrative visualization (Segel and Heer 2010). This is where the narrative visualization is structured linearly and then concludes with an explorative, interactive visualization on the last screen. This structure is quite common for narrative visualization (Segel and Heer 2010).

Visualization B differs from Visualization A as it includes animation. It is based on the American Time Use Survey and is titled "A Day in the Life of Women and Men" by Yau (2019). It is animation-based. It was deemed particularly effective by practitioners due to multiple factors, including the lack of scrolling, the unique animation and the interactive element, which gave control over the animation to the reader. As observed by Segel and Heer, most interactive narrative visualization is a hybrid of both read-driven and author-driven content, which is illustrated succinctly in Visualization B (Segel and Heer 2010).

Visualization C was developed by the authors of this paper as it is not ethical to use a published example as an ineffective example. Published narrative visualization is usually attributed to a particular author, where evaluating it as ineffective would be detrimental to the author. Therefore, to evaluate an ineffective example, it was necessary to independently develop an ineffective example. The ineffective example was based on characteristics of ineffective narrative visualization described by practitioners (Errey et al. 2023). Characteristics of ineffective narrative visualization contained in Visualization C included a complex chart, namely a circu- lar dendrogram, at the opening of the narrative visualization, an inappropriate title, and a lack of cited data source. Visualization C is an example of a 'scrollytelling' narrative visualization (Seyser and Zeiller 2018). Scrollytelling means that through the reader scrolling the story is incrementally revealed in a series of screens. Due to the structured, linear reader experience, it is an author-driven narrative visualization. See Fig. 2 for selected screenshots of each visualization.

### *3.2.3 Participants*

We emailed 48 practitioners of narrative visualization. These practitioners were recruited from an active practitioner forum called the Data Visualization Society. Each practitioner indicated that they contribute to narrative visualization in their personal introduction to the forum. We received 28 responses.

From the 28 that started the survey, a further 11 did not complete the experiment. The final total equaled 17 participants who had completed the entire experiment. Additional details on the challenges of recruiting and requesting time from professionals in narrative visualization can be found in the 'Limitations' Sect. 6 of this paper.

Of those who completed the experiment, four were female, and 13 were male. No participants indicated they were binary/third gender or preferred not to say. We asked participants how many years they had worked with narrative visualization. The largest group was 8 participants, who indicated they had worked with narrative visualization for 10 years or more. We asked participants their primary designation. 5 indicated they worked primarily as developers, 2 as designers, 5 as editors and 5 as 'other.' Some other designations included analyst, consultant and design director. For the remainder of this work, we refer to the 17 participants in the summative study with a P1- ID number.

### *3.2.4 Data analysis*

Summative evaluation is where a published or near finished prototype is evaluated. It is employed in contexts such as industry awards or student assessments. It is important in these contexts that evaluation is

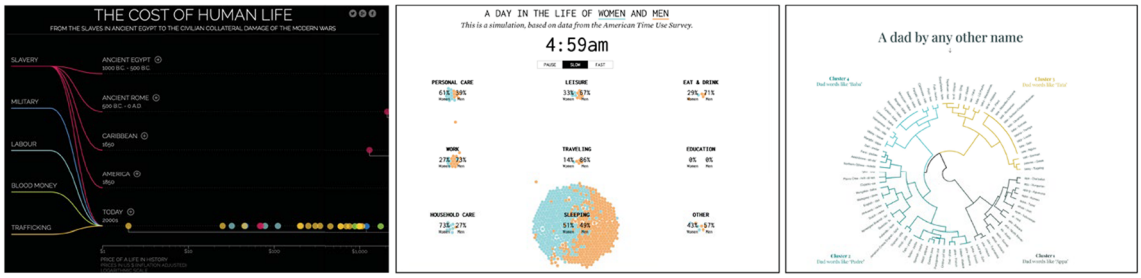


**Fig. 2** Three narrative visualization examples used in the evaluation stage of the summative study. From right, (**A**) by Carlo Zapponi "Has the cost of a human life changed?" (Zapponi 2014), (**B**) by Nathan Yau, "A Day in the Life of Men and Women" (Yau 2019), (**C**) is an example of ineffective visualization devised by the authors of this paper. Permission to feature work was supplied by authors of A) and B)

consistent and repeatable between evaluators. Our chosen method to measure the consistency between evaluators was inter-rater reliability. A high level of inter-rater reliability indicates that the evaluation tool can produce consistent and repeatable results.

Each heuristic was asked in a form of a question, where the practitioner can answer on a 5-point Likert scale. Each point on the Likert scale is correlated with a number, where 1 is strongly disagree and 5 is strongly agree. We compiled the responses for each heuristic. Table 1 shows the mean and the standard deviation for each heuristic, where a rating of 1 means 'strongly disagree,' 3 is a 'neither agree nor disagree', and 5 means 'strongly agree.' For an overall assessment of the heuristic set we aggregated scores. To analyze individual heuristics we focused on individual heuristic scores with a focus on standard deviation. A high standard deviation indicated a lack of agreement among evaluators, where optimally we would expect a low level of standard deviation between heuristic scores.

Finally for open text questions we performed qualitative analysis using thematic coding. We employed a deductive approach, where themes emerged through a bottom- up analysis which is otherwise known as latent coding (Braun and Clarke 2006).

## 3.3 Formative setting study research method

Heuristic evaluation is reported to assist in all phases of development. Our second study aimed to determine if the set of heuristics could aid in the early development of narrative visualization. Generally, in early phases of narrative visualization development low-fidelity wireframes are produced, these support later development, and their audience is usually restricted to the development team. In this study, we investigate if implementing the set of heuristics could improve design. At this early stage, improved design is challenging to quantify, so we relied on feedback from practitioner participants.

### *3.3.1 Procedure*

We ran our two separate experiments with similar procedures. The first experiment was conducted with six practitioners split into two groups. Each group contained three practitioners. We approximated that the skill sets of group members were equally distributed. Therefore, each group consisted of a designer, editor, and data specialist. It has been observed that a group discussion is most effective when group participants have allocated skills to contribute to the discussion (Brilhart and Galanes 1998).

One group was deemed the control group (condition 1). The control group did not use a set of heuristics to evaluate their prescribed narrative visualization task. This group was given 10 min, where the scenario was explained, their expected outcome described and any questions availed. Then, they were allotted 30 min to complete the task described in the task section of this document. The second group followed a similar procedure, including an intervention (condition 2). They were presented with the set of heuristics, which was described in detail for 10 min before their allotted 30 min began. Once the allotted time was complete, we questioned the practitioners on their experiences. Finally, we followed up shortly after the experiment, asking the practitioner to rate their 'satisfaction with their design.'

The second experiment followed a similar procedure to the first. It differed in having only two highly experienced participants and a different task. Similar to experiment 1, experiment 2 gave one participant the set of heuristics (condition 2), and one was deemed the control (condition 1).

#### 3.3.2 Intervention

We adapted the heuristics into a series of questions to implement the set of heuristics as an intervention in our experiment conditions. Their instruction was to implement them as a checklist. See Table 1 for the heuristic questions adapted for the experiment.

#### 3.3.3 Data analysis

We approached the formative setting study with an interpretative lens. Here, our aim was to gain an overall sense of the usefulness of the heuristics from experienced practitioners, rather than to metrically validate them. Our aim was to not generalize the findings of our study rather that the experienced practitioners' advice might be transferable to other practitioners of narrative visualization. Therefore, our data analysis was qualitative, where we employed thematic analysis of interview data using a deductive approach (Braun and Clarke 2006).

#### 3.3.4 Experiment one

*Task* We developed a hypothetical narrative visualization development scenario. The scenario was aimed to resemble a task the practitioners faced in their day-to-day work. We determined an appropriate task by examining previously published visualization work from the participants. The data that the task was based upon is freely avail- able. If the practitioners desired to use imagery, they were encouraged to source the images from iStock photo so that we could purchase their copyright (iStock Photo 2024). The task for experiment 1 (E1) was as follows. A not-for-profit organization named 'Dementia Care Singapore' has requested a narrative visualization that informs Singaporean middle-aged audiences about dementia.

*Participants* We recruited 6 practitioners from a professional data-storytelling studio. The practitioners retained varied levels of experience and expertise. All participants identified as female. For the remainder of this work, we refer to all participants in the formative study with a P2- ID number. Three participants were in the control group that were not given the set of heuristics and three participant were in the group that were given the set of heuristics. See Table 2 for participant characteristics, where Experiment One participants are marked as E1.

#### 3.3.5 Experiment two

*Task* Similar to the first experiment, we devised a hypothetical scenario which resembled a task the practitioners might face in their day-to-day work. The scenario was as follows. An Australian government agency named Cyber.gov.au has requested a narrative visualization that informs an older audience about online scams.

*Participants* We recruited two highly experienced data visualization experts. Each had 11 years of professional experience as data visualization specialists in a range of contexts, including government, finance and health domains. One identified as female and the other identified as male. The male was given the set of heuristics before and the female practitioner was deemed the control, and not given the set of heuristics. See Table 2 for participant characteristics, where Experiment One participants are marked as E2.

## 4 Results

### 4.1 Participant responses

We aggregated the responses of participants to determine the group average and if there were any outliers. Participant responses were generally consistent with the group average. After performing the interquartile range rule, one extreme outlier was determined. P1-4, who submitted the lowest aggregated responses at 83. We did not remove the extreme outlier from the dataset as after analysis, the participant gave valid responses in line with other participants, albeit somewhat lower.

**Table 2** Participant characteristics in the formative setting study

| PID | Condition | Area | Exp. (yrs) | Gender |
|---|---|---|---|---|
| P1 | E1—1 | Design | 6 | F |
| P2 | E1—1 | Editorial | 4 | F |
| P3 | E1—1 | Data Vis | 4 | F |
| P4 | E1—2 | Design | 7 | F |
| P5 | E1—2 | Data Vis | 3 | F |
| P6 | E1—2 | Editorial | 5 | F |
| P7 | E2—2 | Multiple | 11 | M |
| P8 | E2—1 | Multiple | 11 | F |

The first column contains the ID of each participant, the second column corresponds to the experiment they took part in and the condition. The third column contains their area of expertise and the fourth columns has their years of experience. The fifth column contains their gender

The evaluation phase of the survey consisted of 11 heuristic questions per visualization. Therefore, there were 33 evaluation questions in total. The possible aggregated range of responses is 132, where if a participant gave a 1 to each question, the lowest possible score is 33, or if they gave a 5 for each question, the highest possible score is 165. Our reported range of aggregated responses equaled 49; the lowest score was 83, and the highest was 132. For more detail on the inter-rater reliability's significance level, see the following Sect. 4.1.2.

## 4.2 Inter-rater reliability

With the aim of finding out if there is consistency between the responses from practitioners, we tested the inter-rater reliability between scores. We tested inter-rater reliability using the intra-class correlation coefficient (ICC). A high degree of reliability was found between the participant responses. The average measure ICC was 0.88, t(32) = 8.70, p < 0.001. With a confidence interval of 95%0.82–93. The ICC 0.88 indicates a 'good reliability' as the ICC value is between 0.75 and 0.9 (Koo and Li 2016). The 'model' used was the 'Two-Way Random-Effects Model' to generalize results.

This provides participants with similar characteristics, such as professional expertise in contributing to the development of narrative visualization. The 'type' we selected was the 'mean of k raters' due to the multiple participants that took part in the sur- vey. Finally, we chose the 'definition' of 'absolute agreement' as that suited our aim of finding reliable responses to the heuristic questions posed.

We also calculated inter-rater reliability at the individual heuristic level to assess whether the participants' responses were more reliable for some heuristics than others. The analysis revealed that a good level of reliability was found for each heuristic outside of question 7. Question 7 was 'Does the narrative visualization function without the need for interactivity?' and received a 'moderate reliability' score (Koo and Li 2016).

¬Its average measure of ICC was 0.67, $t(32) = 3.16$, $p < 0.056$. See Fig. 3 for the means comparison for each heuristic question per visualization example. This chart illustrates that question 7 had the highest level of deviation and, therefore, displays the lowest level of inter-rater reliability.

## 4.3 Qualitative feedback

### *4.3.1 Feedback on individual heuristics*

In the experiment, we asked participants if they wanted to comment after each heuristic question. It was not mandatory to leave a comment; therefore, they often did not. We analyzed each comment left to gain greater insight into the cognitive processes of the evaluators. We examined question 7 as it received the lowest level of reliability. We found that there may have been some confusion regarding question 7, which asked, 'Does the narrative visualization function without the need for interactivity?' For example, P1-15 wrote in response to Visualization A, "I assume you're referring to the chart page at the end. Interactivity was indeed necessary to advance past the sentence slides at the beginning."

The deliberately ineffective characteristics of narrative visualization contained in Visualization C were reflected in the comments left by participants. Question 2 asked, 'Does the narrative visualization slowly guide the reader into complexity?' We deliberately placed an overly complex chart at the beginning of Visualization C. It seems that the overly complex chart was noticed and assessed as such by participants. P1-

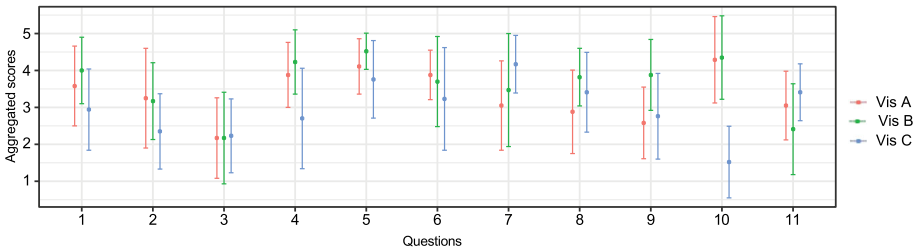


**Fig. 3** Means comparison chart for each heuristic question, grouped by visualization example, with aggregated responses on y axis, bars representing standard deviation per heuristic question

11 commented that Visualization C "Pretty much just jumps right into the complexity, and then the narrative is after." P1-7 commented, "You are given the diagram without much introduction, and it is confusing." This is shown in Fig. 3, where question 2 has a relatively low mean for Visualization C. Similarly, for question 10 relating to data source, it is clear that practitioners recognized that Visualization C did not contain a data source. The relatively low mean for question 10, visualization C is shown in Fig. 3.

We asked if there was any heuristic that the participant would change. P1-8 wrote, "I feel like the question on interactivity could be more nuanced because some of the visuals aren't interactive." This aligns with the finding that question 7 received the lowest degree of reliability. This question asked, 'Does the narrative visualization function without the need for interactivity?' Another comment of interest was from P1-9, who wrote, "The question after data and data source seems somewhat unrelated to the other questions. It's more of a technical/transparency thing. Of course, it is important, but maybe it should be part of a different/technical checklist." P1-11 wrote, "It doesn't really address persuasiveness."

#### *4.3.2 Feedback on the set of heuristics*

The last phase of the survey asked feedback questions. 13 of 17 (76%) participants indicated they 'somewhat agreed' that the heuristics were useful for evaluation. A further two indicated 'they neither agreed nor disagreed' that the set of heuristics was useful, while two 'somewhat disagreed' that the set of heuristics was useful. We asked if the participants would likely use the set of heuristics with complementary guidelines. 12 of 17 (71% ) indicated they 'somewhat agreed' a further four, 'neither agreed nor disagreed.' and one participant 'somewhat disagreed.' We asked an open text question about what other guidelines participants would use in conjunction with the set of heuristics. This question received responses that could be categorized into two major themes. Firstly, responses could be thematically grouped under 'data' as illustrated by P1-5, who suggested a "data focus." Another participant mentioned two specific sets they would use that could be considered data-focused P1-11: "I might use ICE-T or Perceived Visual Informativeness." A second theme that participants mentioned was 'usability' as described by P1-17 "A checklist that focused on visualization best-practice, accessibility, and UX."

The last question on the survey asked participants if they had any final thoughts on the set of heuristics. P1-17 answered, "I love the idea of a checklist to evaluate the efficacy of a narrative visualization, but I feel like this list is too basic and misses some important areas of focus." Conversely, P1-15 responded, "Conceptually, a very good approach homing in on narrative data visualization specifically. Smart list, useful for practitioners who are either actively developing a narrative visualization or critiquing a finished one."

### 4.4 Formative setting study results

#### *4.4.1 Observations*

We asked participants if they found the set of heuristics useful for formative evaluation. P2-7 reported, "It's nice to have something formal rather than just working on intuition." P2-5 explained, "it's a very useful list that we would use. But we would add on other things that might be more applicable." We asked the participants that used the heuristics about their usage process. P2-7, who was part of the second experiment (E2), described that the heuristics functioned as a reference: "Whenever I make a design decision, I can refer back to it and ask which box does it tick?" Those from the first experiment (E1) reported "check through the list to make sure that we had all the things were present" (P2-6). The E1 control group reported

repeatedly discussing the data and re-reading the supplied task. This indicated a less structured approach taken by the control group. Our data suggests that in both experiments, the group that used the heuristic set had a more structured approach to designing and evaluating their wireframe.

While all participants' extensive expertise should be noted, there is the possibility that some heuristics were missed in the design phase. We asked participants if they would have forgotten any heuristic had it not been on the list. Remembering to identify the data source was mentioned in both experiments. (P2-6), who used the heuristic set, explained how they "frantically searched for the data sources, making sure we had them somewhere." It can be observed that in the wireframes designed without the heuristic set, the data source was omitted on their respective wireframes. All participants agreed that identifying a data source is required for a published narrative visualization.

The E1 control group's wireframe concluded with a map, where readers could search for help. In the E2 control wireframe, readers were asked to take part in a quiz. These are both examples of creative ways to engage the reader interactively. We asked why these elements were included in their wireframes. P2-8 explained why she included a quiz, "older people are the audience, and everyone loves a quiz, especially them." We asked the group from E1 why they included a map, "we wanted to say that you can get yourself checked and there are resources to help you" (P2-2). While the wireframe that was heuristically evaluated in E1 did include a quiz-like section, audience consideration was not explicitly mentioned by the group that used the set of heuristics.

## 5 Discussion and future work

### 5.1 Reflections on the assessment

Our analysis of the results suggested that the set of heuristics showed promise for the summative evaluation of narrative visualization. In the summative study, we found that evaluators provided reliable results when evaluating narrative visualization examples. We found a good level of inter-rater reliability when we compared the evaluation responses from evaluators. The intra-class correlation coefficient was .88, where a value between 0.75 and 0.9 is deemed 'good reliability' (Koo and Li 2016). While, there were some discrepancies on an individual heuristic level, the data suggests generally consistent results. When considering that summative evaluation is used in contexts such as industry awards, or student marking, consistent and reliable evaluation metrics mean the heuristic set might prove a useful tool for narrative visualization.

In our formative study we found that the practitioners did find the heuristic set useful. We received a generally positive response to the integration of the heuristics in the design process. We found that the heuristic set provided structure to the design process. The heuristics functioned as a reference where crucial aspects, such as integrating a data source, would be otherwise forgotten if the heuristics were not referred to. One observed drawback of integrating heuristics in the formative stage of development, was that creative audience engagement did not seem to be equally considered by the group that used the heuristic set compared to the control group. More re- search is required to examine if creative audience engagement is less considered when integrating heuristics into the narrative visualization design process.

### 5.2 Complementary guidelines

Practitioners indicated that they would likely use a complementary set of guidelines or heuristics in conjunction with the set studied here. It was clear that the practitioners desired a more 'applicable' set of guidelines or heuristics, which focused on aspects such as usability and data visualization. It is not feasible that one set of heuristics covers all possible aspects of narrative visualization evaluation. Therefore, we have recommended heuristics that might work well in conjunction with the set we have studied.

Usability is widely researched in relation to information visualization. One possible complementary set of usability heuristics, tailored to information visualization, which could potentially translate to narrative visualization, was proposed by Forsell and Johansson (2010) Forsell and Johannson's set of heuristics is a synthesis of multiple heuristics that include, among others, the 10 usability heuristics proposed by Nielsen (1994). Fittingly, Forsell and Johannson's set of heuristics were analyzed and recommended not to be used exclusively, as they lacked aesthetics (Väätäjä et al. 2016).

Another aspect that was not covered in the studied heuristic set but suggested by practitioners was data visualization. Data visualization best practices and guidelines are many and varied. P1-11 mentioned two possible options; "ICE-T or Perceived Visual Informativeness." Firstly, Perceived Visual Informativeness is an evaluation measure that focuses on the quality of visual evidence. Further research is required to determine its suitability as complementary to the set of heuristics we proposed. At this point, it has been studied primarily in the health domain (King et al. 2014). Secondly, ICE-T is a methodology based on a heuristic set adapted into a series of questions similar to the approach we have taken in this work. The data visualization emphasis of the ICE-T methodology means that it could potentially integrate quite favorably with the set of heuristics we have focused on in this work. The primary drawback is that there could potentially be some redundancies, specifically in the upper-level category, the 'Credibility and Trust' category. This category is comparatively data-focused and could be compared to the 'Confidence' category in the ICE-T methodology list of questions.

### 5.3 Refinement of the set of heuristics

We found confusion and misunderstanding concerning question 7 'Does the narrative visualization function without the need for interactivity?' This question received the lowest level of inter-rater reliability. When we investigated the qualitative data, we found further evidence that this question caused confusion. Interaction is a notoriously ambiguous term in visualization research. Retrospectively, we should have adopted an established definition of interaction for visualization, "where interaction is the interplay between a person and a data interface involving a data-related intent (Dimara and Perin 2020)." We believe that citing this definition would have largely cleared the confusion that this question generated. For example, practitioners questioned whether simply scrolling is classed as an interaction, and, as the definition suggests, it is indeed an interaction.

Questions 10 and 11, which come under the upper-level category 'Credibility and Trust', garnered multiple comments from practitioners. Their appropriateness was questioned, as they are more data-focused rather than storytelling-focused. Future iterations of the narrative visualization specific set of heuristics could relate them to a complementary data-focused set of heuristics.

### 5.4 Future research opportunities

A fruitful avenue for future research is further investigation into different complementary sets of heuristics for the evaluation of narrative visualization. The ultimate goal is to determine a suite of heuristics that encompasses multiple aspects of narrative visualization, including data visualization, and usability. A meta-analysis would verify the most suitable usage for each stage of evaluation from formative to summative stages. We predict, that such a study, would greatly benefit the design and evaluation of future narrative visualization.

All participants who took part in this work had professionally contributed to the development of narrative visualization for multiple years. Therefore, the results of this work are indicative of expert advice. It stands to reason that the set of heuristics would influence students of narrative visualization differently. Rather than focusing the attention of an experienced practitioner on the important aspects of the visualization's design, the set of heuristics would act as a means to educate students. A similar usage has been advocated by Santos et. al (Santos et al. 2016). Possible future work could examine if the set of heuristics can foster design skills and provide a useful learning experience for students of narrative visualization.

### 5.5 Exploratory evaluation

The purpose of summative evaluation has been questioned, as it is produced at a stage that is too late to influence the design of the visualization. It is alleged that not much is gained from simply stating one visualization is better than another; rather, an exploratory approach is recommended (Ellis and Dix 2006). Exploratory evaluation asks, 'What have we learned?' Answering such a question should not be limited to a stage in development. The participatory process to the development of our set of heuristics means that with each summative heuristic evaluation, the practitioner acting as evaluator could potentially reflect as to how it could influence their work. We encourage practitioners to share and evaluate the work of their peers. Through an open and supportive dialogue, narrative visualization design and evaluation will be better informed.

In this work, we contribute to the narrative visualization research area by performing a robust and rigorous validation of domain-specific heuristics. Establishing heuristics for specific domains should not stop once the heuristics are proposed. This is an essential step toward establishing a set of heuristics for evaluating narrative visualization; however, it is not the final step. Through an iterative approach, the heuristics will modulate over time to suit this ever-evolving research area.

## 6 Limitations

A common issue with visualization evaluation studies is the dichotomy of localization and generalization. While it is not feasible to study all possible scenarios, we have attempted to use varied scenarios in different stages of development, using topically different materials. For example, we chose formative and summative settings, with different materials. The threat, then, is that the findings in this work are from too varied scenarios to draw concrete conclusions. This is a well-known challenge in evaluation research. We have recognized the fact and compensated, where possible.

The number of participants and the variance in gender are further limitations to this study. The low participant numbers reflect the challenge of finding available professional practitioners of narrative visualization. It should be noted that other work in the visualization research area have similar numbers of participants. Wall et al. recruited 15 participants, and Väätäjä et al. recruited five (Wall et al. 2019; Väätäjä et al. 2016). Our inclusion criteria were quite narrow, and we required a proven record of contributing to the development of narrative visualization. This meant recruiting suitable participants was particularly challenging. Moreover, the gender disparity in both studies was not ideal. Our first study was primarily male participants, whereas the second study consisted primarily of females. Future work will recruit more varied participants and in greater numbers.

Allocating even small blocks of time from busy professional practitioners is problematic. According to practitioner feedback, a lack of time was the primary reason so few completed the survey. At all points in this research, we have considered the time pressures felt by practitioners. We extend our gratitude to the practitioners who have given this research their time and valuable expertise.

## 7 Conclusion

This work presents a step toward empirically validating a domain-specific set of heuristics for the evaluation of narrative visualization. We asked practitioners to evaluate narrative visualizations heuristically, and they produced reliable evaluation metrics. Furthermore, the heuristics were reported to assist in the formative stages of narrative visualization development. We found that the set of heuristics most likely would need to be used in conjunction with data visualization or usability-focused heuristics or guidelines to encompass all aspects of narrative visualization.

**Acknowledgements** We thank all the practitioners who participated in our study. We thank Nathan Yau and Carlo Zapponi for allowing us to use their work.

**Author's contribution** All authors contributed equally to this research.

**Funding** Open Access funding enabled and organized by CAUL and its Member Institutions.

**Data availability** The raw data is freely available to download. See supplemental material provided in Research Method section.

**Declarations**

**Conflict of interest** The authors have no relevant financial or non-financial interests to disclose.

**Consent for Participation** All practitioners agreed to consent form before taking part in research.

**Ethical approval** Ethics approval was provided by the University of Technology Sydney (UTS)

## References

Amar R, Stasko J (2004) A knowledge task-based framework for design and evaluation of information visualization. In: IEEE Symposium on Information Visualiza- tion. Los Alamitos, USA. https://doi.org/10.1109/INFVIS.2004.10

Amini F, Henry Riche N, Lee B, Hurter C, Irani P (2015) Understanding data videos: looking at narrative visualization through the cinematography lens. CHI '15. Seoul, Republic of Korea. https://doi.org/10.1145/2702123.270243

Andrews K (2006) Evaluating Information Visualisations. BELIV'06. Venice, Austria. https://doi.org/10.1145/1168149.1168151

Bach B, Henry Riche N, Carpendale S, Pfister H (2017) The emerging genre of data comics. IEEE Comput Graphics Appl 37(3):6–13. https://doi.org/10.1109/MCG.2017.33

Brath R, Banissi E (2016) Evaluation of visualization by critiques. BELIV '16. Baltimore, USA, pp 19–26. https://doi.org/10.1145/2993901.2993904

Braun V, Clarke V (2006) Using thematic analysis in psychology. Qual Res Psychol 3(2):77–101. https://doi.org/10.1191/1478088706qp063oa

Brilhart JK, Galanes GJ (1998) Decision making in small groups. Effective group discussion. The McGraw-Hill Companies, USA, pp 237–259

Carpendale S (2008) Evaluating information visualizations. In Kerren A, Stasko J, Fekete, North C, Information Visualization Lecture Notes in Computer Science, (Vol Lecture Notes in Computer Science, pp 19-45). Berlin: Springer. https://doi.org/10.1007/978-3-540-70956-5_2

Dimara E, Perin C (2020) What is interaction for data visualization? IEEE Trans Visual Comput Graphics 26(1):119–129. https://doi.org/10.1109/TVCG.2019.2934283

Ellis G, Dix A (2006) An explorative analysis of user evaluation studies in information visualisation. BELIV'06. Venice, Austria. https://doi.org/10.1145/1168149.11681

Errey N, Liang J, Leong TW, Zowghi D (2023) Evaluating narrative visualization: a survey of practitioners. Int J Data Sci Anal. https://doi.org/10.1007/s41060-023-00394-9

Figueiras A (2014) How to tell stories using visualization. IV'14, (pp 18–26). Lis- born, Portugal. https://doi.org/10.1109/IV.2014.78

Forsell C, Johanson J (2010) An heuristic set for evaluation in information visualization. AVI '10. New York, USA. https://doi.org/10.1145/1842993.18430

Hearst MA, Laskowski P, Silva L (2016) Evaluating information visualization via the interplay of heuristic evaluation and question-based scoring. CHI'16. San Jose, USA. https://doi.org/10.1145/2858036.2858280

Hermawati S, Lawson G (2016) Establishing usability heuristics for heuristics evaluation in a specific domain: Is there a consensus? Appl Ergon 56:34–51. https://doi.org/10.1016/j.apergo.2015.11.016

Hullman J, Diakopoulos N (2011) Visualization rhetoric: framing effects in narrative visualization. IEEE Trans Visual Comput Graph 17(12):2231–2240. https://doi.org/10.1109/TVCG.2011.255

Hullman J, Drucker S, Henry Riche N, Lee B, Fisher D, Adar E (2013) A deeper understanding of sequence in narrative visualization. IEEE Trans Visual Comput Graphics 19(12):2406–2415. https://doi.org/10.1109/TVCG.2013.119

Information is Beautiful Awards. Challenges 2014. Retrieved Oct 1, 2024, from https://www.informationisbeautifulawards.com/showcase?type=challenges

iStock Photo. Retrieved October 2, 2024, from https://www.istockphoto.com

Janicki J, Daigle K, Kiyada S, Rao A (2022, December 23) On the Brink. (Reuters) Retrieved October 1, 2024, from https://www.reuters.com/graphics/GLOBAL-ENVIRONMENT/EXTINCT/lbvgggdgevq/

Jeffries R, Miller J, Wharton C, Uyeda K (1991) User interface evaluation in the real world: A comparison of four techniques. CHI '91. New Orleans, USA. https://doi.org/10.1145/108844.108862

King A, Jensen J, Davis L, Carcioppolo N (2014) Perceived visual informativeness (PVI): construct and scale development to assess visual information in printed materials. J Health Commun 19(10):1099–1115. https://doi.org/10.1080/10810730.2013.878004

Koo T, Li M (2016) A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 15(2):155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Kosara R (2007) Visualization Criticism - The Missing Link Between Information Visualization and Art. IV'07. Zurich, Switzerland. https://doi.org/10.1109/IV.2007.130

Kosara R, Mackinlay J (2013) Storytelling: the next step for visualization. Computer 46(5):44–50. https://doi.org/10.1109/MC.2013.36

Lee K, Wong D, Duhalde M (2023) America's debt: what US$ 31 trillion looks like. (South China Morning Post) Retrieved October 1, 2024, from https://multimedia.scmp.com/infographics/news/world/article/3222029/us-debt/index.html

McKenna S, Riche N, Lee B, Boy J, Meye M (2017) Visual Narrative Flow: Exploring Factors Shaping Data Visualization Story Reading Experiences. Eurographics. Barcelona, Spain. https://doi.org/10.1111/cgf.13195

Merčun T (2014) Evaluation of information visualization techniques: Analysing user experience with reaction cards. BELIV '14. Paris, France. https://doi.org/10.1145/2669557.2669565

Nielsen J (1994) Heuristic evaluation. In: Nielsen J, Mack R (eds) Usability inspection methods. John Wiley and Sons, New York, pp 25–16
Nowak S, Bartram L, Schiphorst T (2018) A Micro-Phenomenological Lens for Evaluating Narrative Visualization. BELIV '18. Berlin, Germany. https://doi.org/10.1109/BELIV.2018.8634072
Nussbaumer Knaflic C (2015) Pulling it all together. Storytelling with data: a data visualization guide for business professionals. Wiley, Hoboken, USA, pp 187–205
Qualtrics. Retrieved October 1, 2024, from https://www.qualtrics.com
Ryan M (2004) Will new media produce new narratives? In: Ruppert J, Bernet J (eds) Narrative across media: the languages of storytelling. University of Nebraska Press, Lincoln, Nebraska, pp 337–358
Sadowski C, Kurniawan S (2011) Heuristic evaluation of programming language features: two parallel programming case studies. PLATEAU '11. Portland, USA. https://doi.org/10.1145/2089155.208916
Santos B, Ferreira B, Dias P (2015) Heuristic evaluation in information visualization using three sets of heuristics: an exploratory study. Human-Comput Interact 67(6):259–270. https://doi.org/10.1007/978-3-319-20901-2_24
Santos B, Ferreira B, Dias P (2016) Using heuristic evaluation to foster visualization analysis and design skills. IEEE Comput Graphics Appl 36(1):86–90. https://doi.org/10.1109/MCG.2016.7
Scriven M (1967) The methodology of evaluation. In Tyler R, Gagne R, Scriven M (Eds), Perspectives of curriculum evaluation (pp 39–83). Chicago, USA: Rand McNally
Segel E, Heer J (2010) Narrative visualization: telling stories with data. IEEE Trans Visual Comput Graphics 16(6):1139–1148. https://doi.org/10.1109/TVCG.2010.179
Seyser D, Zeiller M (2018) Scrollytelling – An Analysis of Visual Storytelling in Online Journalism. IV'18. Fisciano, Italy. https://doi.org/10.1109/iV.2018.00075 https://doi.org/10.1109/iV.2018.00075
Tarrell A, Forsell C, Fruhling A, Grinstein G, Borgo R, Scholtz J (2014) Toward visualization-specific heuristic evaluation. BELIV '14. Paris, France. https://doi.org/10.1145/2669557.2669580
Tory M, Möller T (2005) Evaluating visualizations: Do expert reviews work? IEEE Comput Graphics Appl 25(5):8–11. https://doi.org/10.1109/MCG.2005.102
Väätäjä H, Varsaluoma J, Heimonen T, Tiitinen K, Hakulinen J, Turunen M, Ihantola P (2016) Information visualization heuristics in practical expert evaluation. BELIV'16. Baltimore, USA. https://doi.org/10.1145/2993901.2993918
Wall E, Agnihotri M, Matzen L, Divis K, Haass M, Endert A, Stasko J (2019) A heuristic approach to value-driven evaluation of visualizations. IEEE Trans Visual Comput Graph 25(1):491–500. https://doi.org/10.1109/TVCG.2018.2865146
Yau N (2019) A Day in the Life: Women and Men. (Flowing Data) Retrieved October 1, 2024, from https://flowingdata.com/2019/03/06/women-men-timeuse/
Zapponi C (2014) Has the cost of human life changed in history? Retrieved October 1, 2024, from https://humancost.carlo.im
Zuk, T, Schlesier, L, Neumann, P, Hancock, M, Carpendale, S (2006) Heuristics for information visualization evaluation. BELIV '06. New York, USA. https://doi.org/10.1145/1168149.1168162