An Autoethnography on Visualization Literacy: A Wicked Measurement Problem

Lily W. Ge, Anne-Flore Cabouat, Karen Bonilla, Yuan Cui, Yiren Ding, Noëlle Rakotondravony, Mackenzie Michael Creamer, Jasmine Tan Otto, Maryam Hedayati, Bum Chul Kwon, Angela Locoro, Lane Harrison, Petra Isenberg, Michael Correll, and Matthew Kay

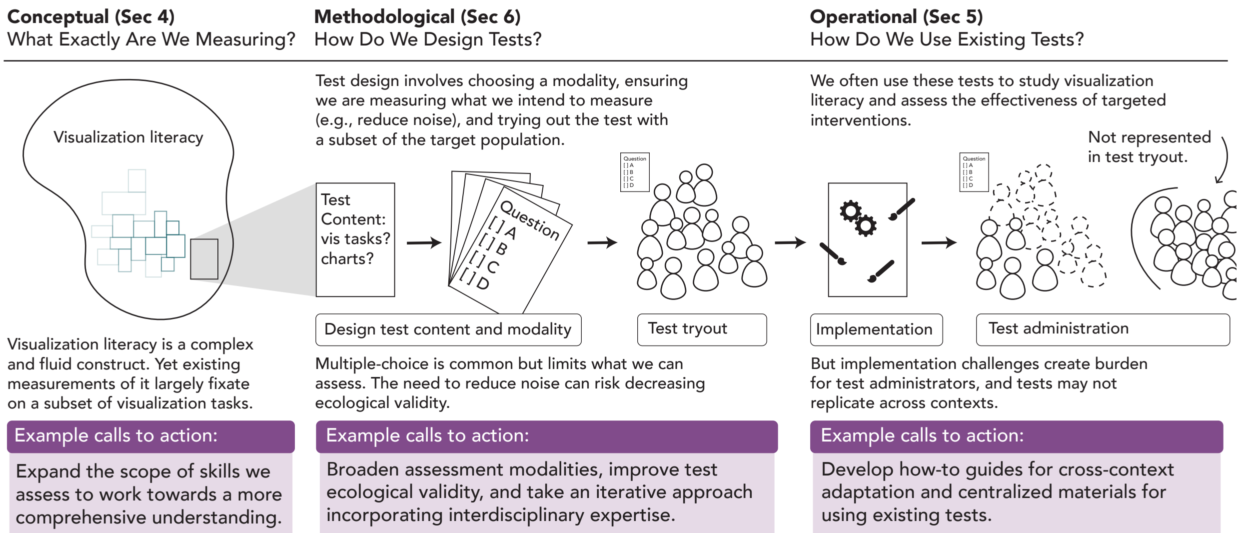

Fig. 1: The lifecycle of a visualization literacy assessment with example calls to action.

Abstract—We contribute an autoethnographic reflection on the complexity of defining and measuring visualization literacy (i.e., the ability to interpret and construct visualizations) to expose our tacit thoughts that often exist in-between polished works and remain unreported in individual research papers. Our work is inspired by the growing number of empirical studies in visualization research that rely on visualization literacy as a basis for developing effective data representations or educational interventions. Researchers have already made various efforts to assess this construct, yet it is often hard to pinpoint either what we *want* to measure or what we *are* effectively measuring. In this autoethnography, we gather insights from 14 internal interviews with researchers who are users or designers of visualization literacy tests. We aim to identify what makes visualization literacy assessment a "wicked" problem. We further reflect on the *fluidity* of visualization literacy and discuss how this property may lead to misalignment between what the construct is and how measurements of it are used or designed. We also examine potential threats to measurement validity from conceptual, operational, and methodological perspectives. Based on our experiences and reflections, we propose several calls to action aimed at tackling the wicked problem of visualization literacy measurement, such as by broadening test scopes and modalities, improving test ecological validity, making it easier to use tests, seeking interdisciplinary collaboration, and drawing from continued dialogue on visualization literacy to expect and be more comfortable with its fluidity.

Index Terms—Visualization literacy, autoethnography, measurement, validity

- *LG, YC, MH, MK are with Northwestern University, U.S. E-mail: {wanqian.ge, charlescui, maryam.hedayati, matthew.kay}@u.northwestern.edu.*
- *AFC is with Université Paris-Saclay, Inria, Institut Polytechnique de Paris, CNRS, LISN, i3, France. E-mail: anne-flore.cabouat@inria.fr.*
- *KB, YD, NR, LH are with Worcester Polytechnic Institute, U.S. E-mail: {kbonilla, yding5, ntrakotondravony, ltharrison}@wpi.edu.*
- *MMC, MC are with Northeastern University, U.S. E-mail: {creamer.m, m.correll}@northeastern.edu.*
- *JO is with UC Santa Cruz, U.S. E-mail: jtotto@ucsc.edu.*
- *BCK is with IBM Research, U.S. E-mail: bumchul.kwon@us.ibm.com.*
- *AL is with Università degli Studi di Brescia, Brescia, Italy. E-mail: angela.locoro@unibs.it.*
- *PI is with Université Paris-Saclay, CNRS, Inria, LISN, France. E-mail: petra.isenberg@inria.fr.*

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx

1 Introduction

A growing number of empirical studies in the data visualization community now rely on measures of *visualization literacy* (i.e., the ability to interpret and construct visualizations) as a basis for designing more effective visualizations (e.g., [40]) and identifying a baseline for skill improvement in participants (e.g., [52]). Prior work has also developed targeted interventions to improve visualization literacy (e.g., [4, 12, 56]), which in turn requires valid measurements for effective evaluation. To meet the demands of such measurement-reliant studies, researchers are thus incentivized to develop—and have already developed—various quantitative assessments of this construct (e.g., [1, 14, 16, 41, 42, 61]).

However, the complex, fuzzy, and multifaceted nature of visualization literacy makes it hard to pinpoint precisely what we are measuring, or indeed *want* to measure. Numerous definitions and frameworks for conceptualizing visualization literacy have been proposed [14, 16, 17, 41, 50, 61, 74]. Different conceptualizations have led to different measurement scales tailored to different visualization literacy components [14, 41, 61]. Initiatives such as meetups, panels, and workshops have tried to distill what it means to be literate in visualizations [43, 54, 55, 72]. Despite these longstanding efforts, debates around what visualization literacy even *is*—let alone how to define it—persist: we still do not comprehensively understand *visualization literacy* as a construct.

Even as we are unable to precisely define it, we must forge ahead, making the compromises necessary to study visualization literacy as both test users and test designers. We view visualization literacy measurement as a *wicked problem*, the kind of problems defined in social planning and public policy as "complex, intractable, open-ended, unpredictable" [3], and which "do not have an enumerable (or an exhaustively describable) set of potential solutions" [73]. The tension between not fully understanding visualization literacy as a construct while still needing to measure it is what makes the study of it so challenging.

We[1] gathered at the ACM CHI 2024 workshop on visualization literacy [43] to begin to tackle this wicked problem. We engaged in critical debates about how to better define, study, and improve visualization literacy. Questions central to our discussions included *what are the assumptions underlying visualization literacy?* and *what worked / didn't work in measurement?* Conversations continued after the workshop, where we reflected in depth on our professional and practical experiences. These reflections exposed our tacit thoughts on visualization literacy—thoughts which often exist only in between polished works and are left unreported in individual research papers.

To bring these tacit thoughts together, we contribute an autoethnography detailing the *fluidity* inherent to visualization literacy—fluidity that makes the construct an ever-moving target across domains, cultures, and contexts of application. As test users and test developers with extensive experience studying visualization literacy, we documented our opinions and reflections in a structured way through internal interviews. Based on a thematic and diffractive analysis of our interview data (Sec. 3), we discuss how the complexity and fluidity of visualization literacy leads to misalignment between what the construct is and how measurements of it are used or designed. From conceptual (Sec. 4), operational (Sec. 5), and methodological (Sec. 6) perspectives, we discuss challenges that threaten measurement validity. From these challenges, we provide several calls to action (Sec. 4.4, Sec. 5.4, Sec. 6.3) that pertain to each of the three perspectives, such as expanding test scopes, making it easier to use tests, broadening test modalities, improving test ecological validity, and taking an iterative approach to test design incorporating interdisciplinary expertise. We hope that by embracing the complexity and fluidity of the construct, the community can better navigate the wicked problem of visualization literacy measurement.

2 Related Work

We outline some of the existing conceptualizations of visualization literacy and assessments and describe collaborative autoethnography and diffractive analysis as methods to surface tacit thoughts.

2.1 Visualization Literacy and Measurement

Literacy, broadly construed, is the ability to read and write [20]. Visualization literacy is often defined analogously as the ability to interpret and construct visualizations [14, 16]. It is an important part of empirical work on improving visualization design. Understanding the visualization literacy of a population can help designers create more targeted, effective visualizations; e.g. health researchers have measured visualization literacy in national populations to understand how to improve visualizations of risk [40]. Measurements of visualization literacy are also needed to evaluate interventions designed to improve it [38], such as educational games for children [1, 4, 13, 39]. Ultimately we can only improve what we can measure, so precise, valid measurements of visualization literacy are widely needed.

However, visualization literacy is not a straightforward construct to measure. The various definitions (e.g., [7, 16, 17]) of it proposed through prior papers, past meetups, workshops, and panels suggest the inherent complexity of the construct (e.g., [43, 54, 55, 72]). Researchers have made many attempts at developing frameworks to conceptualize visualization literacy, such as mapping out the landscape of the construct [74] or reconceptualizing it with visualization competencies, comprehension processes, and practices [50]. There are also individual definitions of visualization literacy focusing on different components of the construct and each leading to corresponding assessment(s). The assessments developed by Boy et al. [14] and VLAT by Lee et al. [61] both measure fundamental visualization interpretation skills (e.g., value comparison). Ge et al. developed CALVI [41] to measure susceptibility to visualization misinformation and AVEC [42] for measuring visual encoding ability in visualization construction. Other works improved the time efficiency of these tests [30, 71]. Other assessments have been designed specifically to evaluate particular interventions [1] or to assess literacy on specific chart types, such as treemaps [36] and parallel coordinates plots [37]. These measurements often exist somewhat in isolation from each other and from efforts in other fields [40, 57, 68].

The theory of *multiliteracies* [21, 25] sheds some light on the complexity and diversity in visualization literacy measurement. That theory describes how literacies can be both multimodal (involving media beyond text [25]) and multicontextual (involving "different cultural, social or domain-specific contexts" [25]). While visualization literacy is clearly multimodal, its multicontextual variability may be what makes defining and measuring it exceptionally hard, leading to the diversity of existing attempts described above.[2] It is in the face of this wicked measurement problem that we looked for a way to both understand the current state of the problem and search for concrete next steps.

2.2 Collaborative Autoethnography

As experts in the area, autoethnography offers a way to document our reflections and tacit thoughts about the state (and future) of visualization literacy—thoughts that are often unreported in individual papers and invisible to literature reviews. Autoethnography is a qualitative research method that focuses on the personal experiences of the authors (*auto*) by describing and systematically analyzing (*graphy*) these reflections for the purpose of understanding cultural experiences (*ethno*) [34, 53]. This method "acknowledges and accommodates subjectivity, emotionality, and the researcher's influence on research, rather than hiding from these matters or assuming they don't exist" [35]. The main source of reliability in autoethnographical work comes from the person's credibility (e.g., could they have had the experiences described?) [35], and a valid autoethnography evokes a feeling that the described experience is "lifelike, believable, and possible" [35].

Prior work in the visualization community has used autoethnography to surface details of complex design processes [31, 64]. However, a weakness of individual autoethnography is that the researcher may be too close to the experience to see it in a nuanced way [58, 59]. *Collaborative autoethnography* builds on top of traditional autoethnography to address this weakness. It involves two or more researchers "pooling their stories to find some commonalities and differences and then wrestling with these stories to discover the meanings of the stories in relation to their sociocultural contexts" [22], and has the potential to create a more rigorous, multi-perspective analysis [45, 58]. Thus we adopt a collaborative autoethnographic approach to gather and analyze our tacit thoughts about visualization literacy.

2.3 Diffractive Analysis

To help us find differences in our pooled stories, we also employed *diffractive analysis*. Diffractive analysis has parallels in feminist and interpretivist approaches that value the differences in researchers' perspectives and view them as important factors in shaping the research analysis [8, 9, 60]. Prior work [2, 60] in HCI and visualization has applied variations of this approach to incorporate analysts' differing opinions into a qualitative analysis. Because one driving motivation for our collaborative autoethnography is to distill the various in-depth experiences that we have accumulated about visualization literacy and its assessment, we sought to avoid passive agreement by employing

[1] A subset of the authors were organizers of the workshop.

[2] We return to multimodality and multicontextuality in Sec. 4.1 and Sec. 4.2.

diffractive analysis to increase the possibility of differences in opinion and encourage constructive disagreements during our analysis.

3 Methods

We lean on our expertise and experiences to seek answers to these motivating questions: *what makes visualization literacy so hard to measure?* and *are existing tests being used as they were intended?*

3.1 Author Backgrounds

The authors of this autoethnography are researchers in information visualization who are either pursuing a doctorate ($N = 8$) or have obtained a doctorate ($N = 7$). We started regular conversations and meetings after the CHI 2024 workshop on visualization literacy [43] that a subset of us organized. While we have all used or considered using assessments of visualization literacy, 9 of us have had experience in developing assessments [19, 29, 30, 41, 42, 49, 61, 65] (Tab. 1).

Table 1: The authors' initials in alphabetical order along with self-reported (at the time of the interviews) years of experience in visualization literacy, visualization research and professional experience, assessment design experience, and other prior work related to visualization literacy.

Author	VisLit (yrs)	Vis (yrs)	Designed Assessment(s)	Other VisLit Work
AFC	1.5	10	[19]	[18]
AL	5	8	[65]	[10, 11]
BCK	9	17	[61]	[43, 54, 56, 62]
JO	1.5	10		[69, 70]
KB	2	2		[32, 43]
LH	9	17	[29, 30]	[12, 32, 43, 47, 66]
LG	3.5	3.5	[29, 30, 41, 42]	[43]
MC	7	15		[26, 27, 43]
MH	4.5	4.5		[43, 50, 51]
MK	7	12	[29, 30, 41, 42]	[32, 43, 50, 51]
MMC	2.5	2.5		[28]
NR	2.5	4		[6, 12]
PI	1.5	20	[19, 49]	[18]
YC	3	3	[29, 30, 41, 42]	[43]
YD	5	5	[29, 30]	[12, 32, 43]

3.2 Internal Interviews

To collect the autoethnographic data in a systematic way, we conducted internal semi-structured interviews with 14 authors.[3] Four of the authors served as interviewers, and each interview was conducted by a pair of interviewers. We coordinated the interview sessions so that the interviewers and interviewees do not have supervisor-supervisee relationships and are not regular co-authors. We asked each interviewee to reflect on how visualization literacy assessments were integrated into their research, their past experiences on the topic, and thoughts on measurements of visualization literacy more generally. The questions we asked during the semi-structured interviews included *what tests do you use/not use and why?*, *what are the issues you've experienced using (e.g., implementing, administering, analyzing) the tests?*, *what do you think visualization literacy even is?*, and *what is your wish list for what you'd like to see covered in tests?*.[4] Asking interviewees to elaborate on the reasoning behind their decisions (e.g., why and how they decided to use certain tests) and to reflect on the challenges they might have encountered along the way allowed us to tap into valuable tacit thoughts on design and use of tests. Each interview session lasted about 45 minutes. We used transcribed audio recordings for analysis.

[3] One author was not interviewed due to time constraints, but they contributed their tacit thoughts during group discussions, analysis, and paper writing.

[4] See supplemental materials for the interview protocol.

3.3 Thematic Analysis and Diffraction

We used thematic analysis [23] to analyze our interview transcripts and supplemented it with diffractive analysis [8]. We employed a series of analysis steps to also embrace differences in opinion and enrich our conversations instead of relying on methods that prioritize agreement among data collectors (e.g., inter-rater reliability [67]). Disagreements among us can offer diverse perspectives to help us form a more holistic view of the many factors that make measuring visualization literacy a wicked problem.

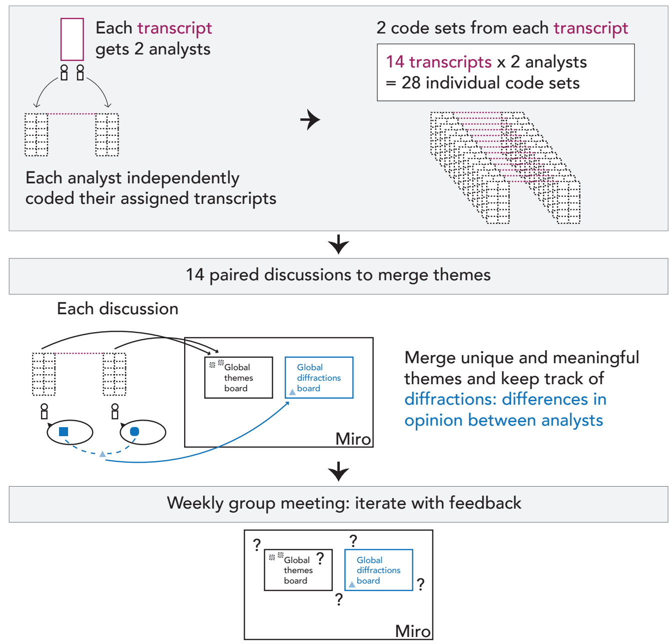

Fig. 2: A flow chart demonstrating our analysis procedure, which involved each analyst individually coding their assigned transcript, pairs of analysts merging their codes and themes on Miro while keeping track of diffractions, and the large group discussing the themes and diffractions.

Independent Coding In total, 9 of the authors served as analysts and followed the same analysis procedure for the 14 interview transcripts (Fig. 2). Each interview transcript received 2 analysts. We started with independent coding to allow the analysts to arrive at their own set of interpretations and codes. To further encourage differences in opinion, we aimed to distance the analysts from the transcripts and made assignments while ensuring the following criteria were met: (1) each analyst did not analyze their own interview transcript, (2) each transcript was assigned at least one analyst that was not one of the interviewers that conducted the interview, (3) each pair of analysts was not from the same research lab, and (4) each analyst did not analyze a transcript from someone from their research lab. Each analyst independently coded emerging themes from their assigned transcripts.

Paired Discussions Once both analysts from one transcript finished their independent coding, they met to discuss their codes. During their discussion, they merged their independent codes into theme sets while keeping track of differences in opinion between them (i.e., diffractions). This process allowed analysts to reflect on their interpretations, compare and contrast their understandings, and take note of any diffractions that emerged. The analysts contributed to two Miro boards during their discussions.[5] One Miro board documented the emerging themes and sub-themes (the global themes board), and another Miro board was dedicated to any diffractions that surfaced during the analyst discussions (the global diffractions board). All of the analysts edited and / or added to the same two Miro boards, meaning that the themes from different analyst pairs were merged as different paired discussions happened. Two types of diffractions emerged: (1) disagreements within a pair of analysts on interpretation of particular quotes, and (2) disagreements with another pair of analysts on their code(s) or theme(s).

[5] See supplemental materials for the Miro boards.

Large Group Meetings We held weekly group meetings that included authors who were not analysts to discuss the global themes board and global diffractions board. This allowed us to incorporate non-analysts' perspectives and to surface differing opinions and refine themes. Discussions of diffractions led to (1) resolving disagreements on interpretation of particular quotes by asking the interviewee of the quote to clarify and (2) large-group discussions about any new themes and how to better structure existing themes. The diffractions served as the main discussion points during these meetings, guiding the analysis.

Themes emerging from our conversations revolved around conceptual (Sec. 4), operational (Sec. 5), and methodological (Sec. 6) challenges. We detail these challenges and how they may lead to misalignment between what the construct of visualization literacy is and how measures of it are used and designed.

4 Conceptual: What Exactly are We Measuring?

A challenge we raised repeatedly throughout our interviews was the difficulty of pinpointing exactly what visualization literacy is. Having a broad definition, such as "the ability to interpret and construct visualizations", may be helpful in communicating a general idea about the topic, but it is less useful when the goal is to operationalize it and design a precise measure of it. Here, we describe the factors that contribute to the complexity and fluidity of the construct, thereby making it difficult to define and conceptualize. We also discuss the term *visualization literacy* itself and our diverging opinions on its usefulness.

4.1 Sources of Definitional Complexity

Visualization literacy is a complex construct, yet definitions of it often seem overly simplified or do not fully capture what we think might constitute being "literate" in visualizations. This feeling of inadequacy is partly due to the many implicit and often hard-to-describe skills that contribute to making someone proficient in visualization. **Visualization abilities draw on knowledge from a variety of disciplines**, such as mathematics, data science, statistics, computer science, geography, design, art, cognitive science (to name a few), as many of us pointed out when asked *what do you think visualization literacy even is?* This complexity echoed the sentiment towards visualization from prior literature [7] and multimodality within the theory of multiliteracies [21, 25]. As BCK stated:

> Visualization literacy [...] consists of multiple skills, multiple skills in different levels, in different depths. 'cause I mean, math I think is involved in it. Some sort of sense-making in the actual literacy of understanding human language is also part of it. So everything is intertwined—graphicacy, understanding shapes, geometries, altogether. [...] To be able to use the visualization to solve problems, I think you need to understand the construction skills, understanding the mapping between data and visualizations. (BCK)

Indeed, cognitive skills, such as spatial reasoning and verbal processing, along with sensory abilities like color vision, also interact and play a crucial role in applying visualization skills [63], making the construct even more difficult to isolate. But we should embrace its interconnectedness:

> The community probably should go [toward] understanding where visualization literacy fits along with other cognitive characteristics, traits, skills, and how much they are influencing each other. (BCK)

In addition to skills from related domains, there are also knowledge and skills specific to data visualization, such as visualization grammar, graphical conventions, and visual metaphors. Research efforts to categorize visualization skills in varying granularity have led to a vast collection of visualization task taxonomies (e.g., [5, 15]). However, **we were broadly dissatisfied with tasks typically used to define visualization literacy**, such as low-level tasks like value retrieval or data comparisons:

> Things like retrieve value is one of the common tasks in these visualization literacy assessments. But [...] if I want to retrieve a value, I go to a table. (YC)

> Like a Cleveland and McGill [task]: estimate the ratio between two values in a chart. Like, that's just not a thing that people seem to need to do. (MC)

While there is value in understanding whether people can read data from a given chart, there are other higher-level (and perhaps fuzzier) visualization skills that are just as important (if not more) to assess:

> A more complete idea of what visualization literacy should be [...] more on the side of, what do we actually use visualization for? And not just can you read a bar chart. (MK)

We need to improve the ecological validity of visualization literacy assessments by focusing on skills that are representative of real-world use cases, such as decision-making, the ability to infer new knowledge from a visualization, or the creation of effective[6] visualizations. Although there is nascent work assessing such skills [1, 42], there is still an inadequate representation of these higher-level tasks in most widely-used tests. If this imbalance persists, it will limit our ability to reach a more comprehensive understanding of visualization literacy.

The complexity of the different skills and knowledge that make up this construct makes it difficult to precisely define and conceptualize. **It is difficult to measure what we cannot fully describe.** One source of this definitional difficulty may come from differences in how people learn about visualization, which may be through exposure (e.g., implicit internalization of visualization rules by viewing real-world examples) or through more formal settings like classrooms. Thus, different people may have different mental models of how visualizations work. By analogy to language, MK explained:

> I think most people, when they're speaking their first language, are probably not as good at the technical grammar of that language as people who are second language speakers sometimes. Because as a second language speaker, you often learn those specific rules of grammar, right? And as a first language speaker, you internalize them, but unless you actually sit down and learn them, you often can't articulate them. (MK)

Two people who are equally able to read, write, and speak a given language may differ in their explicit knowledge of lower-level skills, such as grammar. Similarly, the specific skills underlying a proficiency in data visualization and the explicitness of those skills might differ from one person to another, depending on how they learned. These differences contribute to the difficulty of identifying a precise set of skills that can define what it means for someone to be literate in visualizations.

4.2 It Changes Across Domains, Cultures, and Time

Not only is visualization literacy complex, it is also quite context-dependent, a property we refer to as its *fluidity*.[7] For instance, many examples from our conversations highlighted that **visualization literacy is domain-specific**. The skills required to understand visualizations in a medical or healthcare context may be vastly different than what is necessary in business intelligence. Early work in visualization literacy assessment development also pointed out that domain knowledge could play a role when applying visualization skills [14]. This raises the questions of if it makes sense to attempt to define *visualization literacy* as an immutable construct across domains. Perhaps this domain-dependent characteristic resembles language *dialects*:

> You have academic visualization literacy or [...] business intelligence visualization literacy or whatever. I think that those are different dialects. (MK)

If someone is "literate" in one specific visualization domain (e.g., business intelligence), that does not imply that they are immediately "literate" in another (e.g., data journalism). JO recalled:

> I've heard these fun anecdotes about [...] "oh, when I make charts for these business people, I have to use this particular kind of chart. If I present the same information in a different chart type, they don't care, right? It doesn't look right." (JO)

We considered if perhaps, amid different visualization dialects, it would be useful to look for a *minimum standard* for what constitutes someone being "literate" in visualizations, or what a minimum standard might

[6]Though what "effective" means also varies across contexts (Sec. 4.2), making these higher-level skills even fuzzier.

[7]*cf.* the multi-contextual dimension of the theory of multiliteracies [21, 25].

mean in particular contexts. There were many diverging opinions: a minimum standard might include basic skills to get through life (e.g., map navigation skills), the skills needed to ensure the quality of information consumed (e.g., critical thinking), or even an understanding of fundamental syntactic elements of visualizations (e.g., axes and scales). While we could not form consensus, these might serve as starting points for identifying what we need to measure within different contexts.

The fluidity of visualization literacy does not stop at the domain-specific level. As alluded to in Sec. 4.1, how people develop visualization skills may differ due to their cultural background and experiences: **visualization literacy is also culturally-dependent**. This dependency leads to differences in what people consider to be important in a visualization or what they use visualizations for. A simple visualization reading task implicitly requires culture-dependent knowledge such as its language, local maps, or region-specific political information. As MC put it, visualization literacy "is also this very contingent and culturally-bound thing." By analogy to language literacy:

> You can be very literate in English without having read a Shakespeare play [. . . So] it's gonna be hard to make a universalizable principle [about visualization literacy] because so much is bound up with sort of the visual and graphical cultures that are associated with the types of visualizations that people will need to interpret. (MC)

Any attempt to define visualization literacy as a static construct that is absolute across cultural contexts would be an oversimplification, potentially compromising future assessments and interventions.

The fluidity of visualization literacy is amplified if we step back and consider it on a larger time scale. A few of us (BCK, LH, MC, MK) who have been studying visualization literacy for a longer period of time reflected on how attitudes in the community have changed over the years, suggesting that **visualization literacy also evolves over time**. Since the early conversations about the construct, the conceptualization of visualization skills has shifted from a definite set of skills, stemming from precisely-delimited visualization taxonomies, to a more complex understanding of visualization literacy involving skills and cognitive abilities from other disciplines, such as numeracy and spatial reasoning. BCK reflected on years of experience working in visualization literacy:

> My understanding of visualization literacy has evolved. After a series of follow-up studies, conversations, and studies conducted by other people, and also by just pure observations in the field [. . .] After 10 years, I think the visualization literacy [. . .][skills are] all combined in interesting ways [. . .] using [visualizations] to solve different problems require different sets of skills. And it's also highly contextual, I think. So in certain cases, for a certain context, you may have a higher chance to solve the problem with a given chart. Maybe in other context you don't necessarily. (BCK)

Moreover, some of us anticipate technological advancements changing the way we interact with or create visualizations, which could affect the conceptualization of the construct and subsequent assessments. KB wondered, "how do you even disentangle technical skill from actual visualization skill?" While technical skills play a part in creating visualizations, several of us do not see technical skills as a necessary part of visualization literacy for a general public, as MK expressed:

> If you couldn't make [visualizations] in like ggplot or matplotlib or D3, that's fine, because most people don't have to do that. (MK)

In that respect, perhaps technological advancements would not have a large impact on what constitutes visualization literacy. YD also felt visualization literacy will not become obsolete:

> Right now the AI probably can summarize the information from the charts. [. . . But] I think people with good data visualization literacy can definitely retrieve the information they want from the charts much faster than [by asking] the AI. (YD)

As researchers in visualization, it is in our best interest to be attentive to how visualization literacy differs across different contexts (e.g., domain, culture, and time) and to stay open to the possibility of change.

4.3 Differing Opinions on "Visualization Literacy"

The term "visualization literacy" itself is often the source of first impressions about the construct. This is especially important when we are communicating within our own research community but also when disseminating research output to other fields. We had diverging opinions on the utility of the term "visualization literacy": while some of us saw it as a useful metaphor, others advocated for finding other terms.

"Visualization literacy" draws an explicit parallel to language literacy, which has some advantages from an empirical standpoint, such as how we might conceptualize the construct in a way that is useful for measurement:

> Maybe we should be borrowing more from the literacy metaphor in terms of thinking [of] it as this longitudinal lifelong thing with multiple levels that you're doing as opposed to a psych assay with some low to high point scale. (MC)

Other than using it as a metaphor to learn from the language literacy domain, the term was less attractive for some of us. MH, for example, is hesitant to use the term when talking about specific skills:

> I try not to use the word "literacy" because I feel like literacy has so many connotations that come with it in terms of societal implications and in terms of when you were calling someone literate, you were also calling someone else illiterate. (MH)

This echoes a prior sentiment from MC [27], which he elaborated on during his interview when asked, *have you encountered any challenges communicating visualization literacy as a concept?*

> The main challenge as mentioned is that I don't think it's a meaningful concept. But I've definitely already lost that battle with respect to the vis community, right? [. . .] I feel like [the visualization literacy concept] is often used as an excuse for not doing a good job as a designer, or often as a way to maybe downplay effects that we see [. . .] that challenge conventional wisdom about data visualization. Like, they'll go [. . .] "well, you see this thing in graphical perception, but that's probably with, you know, low graphical literacy crowd workers, of course doesn't apply to us." (MC)

When visualization literacy guides how we design and evaluate visualizations, the "literate" versus "illiterate" split that the term imposes may not serve the goal of improving design for a wider audience [27]. There are also inconsistencies associated with the usage, because visualization literacy can encompass many different complex skills (see Sec. 4.1):

> If you say the word "literacy" to 10 different people, there's gonna be 10 different kinds of internal understandings of what that is. (MH)

Some of us worry that **the broadness of the term and its inconsistent usage reduces clarity**. Those who expressed this sentiment proposed alternative terms such as "graphical comprehension", "visualization ability", "visualization competency". LH differed:

> I think people will know what you mean when you say "visualization literacy", but if the term were changed to, I don't know, just "visualization ability", I think it would potentially fall flat and maybe fail to connect with the audiences outside of visualization that we wish to engage with. So I'm a little comfortable with ambiguity and impreciseness in service of connection. (LH)

A more broadly-familiar term might be an acceptable compromise to ease conversations with other related domains (e.g., statistics, cognitive science). Still, our differing points of view on the term itself further reflect the complexity of visualization literacy and our varying priorities.

4.4 Calls to Action

It is difficult to pinpoint exactly what we are measuring, as visualization literacy is an inherently complex, context-dependent construct. This poses definitional challenges, which inevitably lead to measurement challenges. To progress, we must acknowledge we are tackling a wicked problem and be ready to make compromises in definition and measurement.

Expand the Scope of Visualization Skills We Assess

We generally expressed concerns about the limited scope of visualization tasks measured in existing assessment tests and a need to improve the ecological validity of visualization assessments (Sec. 4.1). Future work should focus on assessing higher-level activities with data visualization, such as decision-making or the ability to create data visualizations. We should **seek ecological validity by focusing tests on what people actually *need* to use visualizations for**. Although the community is already making efforts to shift the focus towards high-level tasks like visualization construction [1, 42] or the detection of misleading representations of data [41], there is still much to do to expand the scope of higher-level skills we assess. Other opportunities to broaden the scope of assessments include measures of skills needed to make sense of different types of data displays (e.g., interaction, animations). Indeed, existing tests mostly rely on static representations, while many visual displays of data are interactive.

Current tests also typically only cover common, basic chart types (e.g., cartesian planes and maps). We should **expand the scope of assessments to less common—but useful—chart types and idioms**, such as icon arrays, which deviate from what we can expect in a K–12 curriculum. Given the difficulty of precisely describing the underlying skills involved in visualization proficiency (Sec. 4.1), we may also lean more on qualitative methods to surface implicit visualization skills before attempting to assess them. Broadening the visualization skills we assess, however, will require careful consideration of potentially-complex relationships between those skills in order to ensure valid measurement. For example, a person's ability to understand interactive visual data displays might partially depend on their ability to interact with digital interfaces at large. While we offer some ideas for broadening test scopes, it would be a useful exercise to speculate about what the world of visualization literacy measurement could look like, as a way to inform how it should be.

5 Operational: How Do We Use Existing Tests?

Despite the difficulties with conceptualizing visualization literacy, researchers have already attempted to measure it to better understand their target population's visualization literacy levels [40] or evaluate the effectiveness of interventions to improve visualization skills [4, 13, 39]. As test users, we share the considerations that go into deciding when and how to use a particular test, such as our research goals and community norms. We also reflect on the assumptions that often come with these tests, which can influence our decisions on test usage in practice.

5.1 In Pursuit of Standard and Objective Measures

Researchers use existing visualization literacy assessments for a variety of reasons. One reason that emerged is **to measure visualization literacy in an objective way**. We note that early assessments answered a then-emerging need in the community for more objective measures:

> Before we had the assessment test, we might ask people to self-report their comfort with things like statistics or visualizations. And I remember distinctly from that that some of the times people greatly overestimated or they put like a really high score on their self report, and yet the performance was pretty low in comparison to other people that we tested. (LH)

The desire for more objective tests in the community led to the wide adoption of some of these early assessments, as LH reflected: "they were kind of widely adopted, because people were literally searching for this sort of thing and it was available" . Because VLAT, as one of the earliest tests in the community, offered a more objective way to assess visualization skills than what was previously possible, it made sense that most of us who used tests in experimental contexts have used or considered using VLAT. LH explains that part of this test's success stems from its foundation in existing community standards:

> VLAT is very logically constructed in that it kind of builds on well-accepted task taxonomies [...] So, there's like a logic to it that's very easy to explain and very easily could result in like a series of questions. And then the scoring is very logical as well. (LH)

Across our interviews and conversations, there were many instances where we described choosing the latest community standard when making decisions about how we conduct our studies. Reflecting on these matters, we note that **community norms also exert influence on study design choices**. These observations reflect a potential desire in the community to have a "universal" test to produce "generalizable" results. Having a standardized measure becomes an advantage in research contexts when the goal is to compare results across multiple studies. In such cases, we might choose the same measurement scale as in existing work. However, MC pointed out the tension between efforts to obtain generalizable results and the need for scales tailored to specific study interests:

> If everyone's using their own metric, then it's hard to make these generalizable or universal claims, but if really what we want to get out of those scales are very-specific-skills-based, or are people reliably pulling the right things out of this [specific] chart, then that sounds to be much more contingent and less universal. (MC)

We as researchers should carefully consider such trade-offs and be clear on the goal of our studies when selecting a measurement scale, as **the most appropriate measure might not be the community standard**. We acknowledge that there is no easy solution to resolving this tension. On one hand, an appropriate test that is well-suited to a particular research question may not always exist, often due to the difficulty of conceptualizing what we want to measure (Sec. 4). On the other hand, when an appropriate test is not one that is a standard in the community, community norms can impose pressure on researchers to not use it.

Our discussion also revealed a need for assessments that extend outside of the research community. For example, some of us have received outreach from industry about using VLAT "as a part of a manual or tutorial for their company workshop or orientation" (BCK), or A-VLAT [30] (adaptive VLAT) to assess the visualization literacy of a company's employees. This reflects a desire in industry to use these assessments as "[part] of their onboarding process, onboarding materials" (YD).

Interest outside of academia demonstrates the potential for increased impact of visualization literacy assessments, yet also highlights further limitations in the current assessment landscape. For example, current assessments are typically designed with no feedback, and assign a single number (score) only at the end. This makes individual or group learning difficult or impossible in some cases, and highlights barriers that would need to be addressed before organizations could make effective use of the current set of visualization literacy assessments.

5.2 Implementation Challenges Burdening Researchers

Reflecting on our practical experiences using tests, **it was a lot easier to adopt tests with readily available materials**, which also potentially contributed to the wide adoption of VLAT: the straightforwardness of implementing and scoring VLAT made it easy for researchers to use. However, we also surfaced several practical challenges during use that influenced our decisions in choosing tests. One of the most often mentioned barriers was that the **test length may be too long for some study designs**, which could introduce an unnecessary burden to participants. Researchers often evaluate how a test might fit into their overall study design:

> If you are just trying to check their experience levels, it feels very costly for the users 'cause [it] takes an hour or 30 minutes to an hour overhead cost and your main test might take three minutes. (BCK)

The time requirement for some of these tests led to efforts to reduce the length, such as shortening it [71] or adaptively and more efficiently assessing the ability [30]. **Some of these enhanced methods require more technical knowledge to implement**, making them harder to adopt:

> I think there are a lot of potential technical hurdles, like little engineering things that don't seem like a big problem, but could really affect the choices that a research team makes, especially under time pressure. [...] Running the psychometric or statistical models to analyze participants' data on the assessments requires some knowledge, first of all, theoretical knowledge about how those models work, but also more practical implementation knowledge. [...] We were lucky in that we received a lot of help and from people with these expertise. But generally speaking, I think it's not a trivial process. (YC)

5.3 Non-replication of Tests Across Contexts

Because each test was created within a certain context, it likely comes with assumptions about the test population. Such assumptions relate to some baseline knowledge that the test-takers are expected to possess regardless of their visualization skills, such as understanding a language (e.g., English) or being familiar with the topic of a visualization (e.g., geospatial maps, politics), or knowledge specific to their educational level (e.g., high school, college). **Cultural assumptions impact how researchers adopt the tests in differing contexts.** As existing tests were originally developed in English, with Western-centered cultural assumptions, there is a need for translating and adapting these tools to other languages. BCK recalls that "there were at least two or three people asking me for permission to translate [VLAT] into their own language." However, adapting a measuring instrument across cultures is a time-consuming process involving terminology challenges and revalidation steps such as back-translation. NR, who was motivated by the cultural dependency of the visualization literacy tests like VLAT and worked on cross-cultural adaptation, recalled many trade-offs:

> We absolutely did not know how do we say "heat map" in [Arabic] and we just made a compromise of something that closely sounds like a heat map if someone heard about it. So that was a challenge to find the right translation, the taxonomy, in the other language. (NR)

Not only did they need to translate the language used in the tests, NR also had to modify the topics of visualizations and "replace the question with local election systems". This was necessary to be mindful of sensitive topics across different cultures. Sometimes, the topic of a visual representation could even derail test administration and defeat the purpose of evaluating visualization skills. NR recalled:

> When we tried to ask them like, here is a chart about transportation, I think it was about the metro [...] and we asked them like, tell us what you learned from this chart. And they just ended up [...] talking about how messy transportation is in Madagascar [...]. If the culture of reading a chart is not yet established, then people can see the chart as a trigger for parallel discussions, [which] is not necessarily relevant for the evaluation at the moment. (NR)

This *non-replication* across cultures challenges our rationale in Sec. 5.1 to use the same test to compare results across studies, suggesting that it may only be possible to have "standardized" measuring tools for comparisons within a particular context of use. Using *dialects* as a metaphor for visualization contexts, MK noted: "A lot of quote unquote rules of visualization are actually rules that are rooted in specific dialects." Such differing contexts are not limited to cultural differences, but may also concern the application domain, such as statistical or spatial data. For example, MMC, who has worked with astronomy visualizations, found a lot of the current assessments to be not particularly useful:

> A lot of what we look at for visualization tests measures one's ability to infer statistical data, which maybe perhaps isn't necessarily as helpful as like scientific data with seeing like a galaxy formation or trying to come up with a nebula or anything like that. (MMC)

Current tests can only represent a small snapshot of the construct under a very specific context, and may not be built to accommodate the fluid nature of visualization literacy. We should therefore communicate these domain-specific and culturally-dependent test assumptions more clearly to facilitate future test usage.

5.4 Calls to Action

How we choose to use a particular visualization literacy test is influenced by factors like research goals, community norms, and ease of implementation. The unspoken test assumptions regarding population characteristics can also lead to usage barriers. We call on test designers to make tests easier to use and the community to build centralized repositories.

Develop How-to Guides for Customization and Adaption

Many of us wished to adapt existing tests while maintaining construct validity. Because of the rigorous test design and validation processes (Sec. 6.2), later test administrators might feel reluctant to adapt tests in fear of affecting the test validity. Yet in some cases, alterations of questions may be more desirable in a particular study context (e.g., cross-cultural adaption in Sec. 5.3 or shorter studies in Sec. 5.2) that is different than that of the original test questions. **It is crucial to clearly communicate test assumptions and the knowledge necessary for adaptation**, so users can better adjust the test in different contexts. LH entertained the idea of having a "how-to" guide for customization:

> I think that the main limiting factor [is the absence of] a widely accepted way of paring down the questions or slicing and dicing, and what that means if you do that. I would like to have the paper that says "this is justified', or "this is how you construct something for a particular context" with justification [...] I would follow that process potentially, if it wasn't too onerous.(LH)

This could help reduce the burden researchers might feel when they need to make alterations, and could improve the alignment between study designs and measurement.

Form Centralized Assessments and Implementation Materials

The availability of materials is a huge factor in implementing and administering tests (Sec. 5.2). **It is crucial to make tests and the necessary materials readily available.** We can also benefit from a collective effort toward a centralized repository that contains the different assessments, including a modular approach to cover the multidimensionality of the construct. Such a centralized approach could preserve the benefit of having individual tests for isolating particular components of visualization literacy for measurement and, at the same time, offer easily accessible ways for users to find and implement their desired tests. Existing efforts in the community are already taking steps toward better assisting users in implementing and administering tests, such as the reVISit framework [33]. We can build upon and extend prior efforts in centralizing these tests to lower the barrier for using assessments of visualization literacy.

6 Methodological: How Do We Design Tests?

Designing a valid test is time- and resource-consuming. The subset of us who have designed and validated existing tests all acknowledged that it was no easy feat. Many considerations went into designing valid tests, such as what we want to (or can) measure and how we can measure it. Depending on what we are measuring, the way we assess the validity of a test also differs. There are many challenges during test development from a test content and format perspective, which lead to the compromises we often have to make as designers of visualization literacy assessments.

6.1 Test Modality Limits What We Can Measure

As a community, we took very logical first steps in studying and measuring visualization literacy by seeking theoretical and methodological insights from the psychology and education sciences, such as Item Response Theory [14] and systematic test development frameworks [24,61]. These processes remain important sources of structure for test design. However, **areas of test design such as test format and scoring have stayed relatively constant over time**, despite the fluidity of visualization literacy as a construct. This misalignment impacts what we can measure with existing test formats.

We observed a tendency for test designers to lean more on quantitative measures. This tendency then favors test modalities that are easy to implement and summarize quantitative results from, such as multiple choice questions.

> There is a general assumption that if you're gonna build an assessment, you want to build something where you don't have to have a human sit down and grade it. Which, you know, may or may not be a good baseline assumption. (MK)

While multiple choice questions can provide a range of benefits from an implementation perspective, such as being easy to grade, there are still challenges during test design. One challenge is **designing meaningful distractors in addition to not-so-obvious correct answers**. The quality of distractors in multiple choice questions has a direct impact on the question difficulty and test validity at large. For instance, if the correct answer is too obvious, then the question may not capture the visualization skill it aims to assess. Depending on the purpose of the test, sometimes distractors need to be designed in a very specific way, as LG recalled:

> For CALVI, [...] there is definitely an answer that is considered correct, but also [a distractor] that is seemingly correct if they didn't realize the issue. So that makes it a little bit iterative because we needed to make sure that the visualization for that question can actually mislead the person, if they didn't notice that issue, to choose that answer. We had to think about [the visualization and the distractor] together when designing the question, the visualization, and the misleading answer. So that was a lot to consider at once, but for each item, that was necessary because we needed something to catch whether they identified the issue or not. (LG)

The challenge of designing meaningful distractors may be easier to overcome for measurements of visualization interpretation skills compared to construction-related skills, as the *combinatorial* nature of visualization construction is less likely to be supported by the multiple choice format [42]. **The multiple choice modality thus limits what visualization skills we can assess.**

Following the multiple choice questions format, the outcomes of these tests naturally become aggregated scores—sometimes broken down by chart type. This straightforwardness in scoring also contributed to the wide adoption of some tests, as discussed in Sec. 5.1. However, because the results output, or scoring, influences the amount of information researchers can retrieve from using assessments, **the simplicity of the resulting score may lead to less-refined analyses of visualization literacy as a construct**. Noticing this limitation, several of us voiced a desire for more insight on test-takers' visualization abilities and suggested leaning more on qualitative methods.

> I wish we could learn more about how people get things wrong when they take the assessments [...]. Right now it's mostly like a number, so that's less meaningful. So maybe more open-ended questions on these assessments. But of course that introduces additional complications such as grading and not being objective. But that is the trade-off. (YC)

> I would love to see scoring rubrics for [...] open-ended answers, which are constructed responses [in learning sciences]. They come with this "rubric" that helps you score analytically or holistically. (AFC)

Offering a different reason on why we should adopt more qualitative rather than quantitative ways of assessment, MC expressed:

> There's a lot of things that we seem to assess quantitatively in the vis field that I don't know if we've done the necessary qualitative observational stuff that seems to be a prerequisite for some of those things yet. [...] I just don't know how people read bar charts, for instance [...] Are they anchoring at the tip of that line and then projecting it to that y-axis and then converting that to a number in their head, and then doing comparisons of that number? Are they doing [...] very quick area estimations? [...] So when you get from there to "our design was better because they completed task one, like 0.5 seconds faster on average", there's a huge gap between those empirical claims to me and how useful those empirical claims are, or quantitative empirical claims. (MC)

This relates to the implicit knowledge that we might internalize through experience (Sec. 4.1), which could be key to interpreting some of the results from quantitative assessments. Surfacing this implicit knowledge, such as through qualitative methods, might be necessary to lay a solid foundation for building quantitative assessments. However, the very reason that contributed to the popularity of these tests—logical test structure and scoring—may now be driving people away from alternative test formats that contradict those properties. LG wondered:

> Are we just using multiple choice as a default without thinking about the more complex ways of measuring stuff? And if we are, then we are maybe missing some opportunities to be able to measure these higher-level skills. (LG)

While we were able to rely on certain test formats like multiple choice questions to collectively gain valuable insights on visualization literacy, staying too comfortable with current attempts of assessing the construct may lead to missed opportunities for eliciting and understanding all of the exciting aspects of visualization literacy, as choices on test modality also directly influence what we can effectively measure.

6.2 The Need to Reduce Noise

When designing test questions, it is difficult to determine the correct level of abstraction to operate on. In other words, the granularity of the categorization of visualization skills directly impacts what each test question will focus on. Test designers often have to determine the taxonomy of visualization tasks to base their test questions on, such as lower-level or high-level categorizations of visualization skills. A more granular categorization could result in redundant test questions and risk the diversity of the question bank. A less granular categorization, however, might lead to difficulties in translating abstract categorizations into realistic and concrete test questions. As MK reflected,

> You can create this whole abstraction, this nice taxonomy, you think that it works well and then you start making the actual charts and you realize [...] it's more difficult when you actually see the real thing. (MK)

This creates a challenge that test designers often have to battle with, which first starts with some abstractions (e.g., visualization tasks), and "the question then becomes, okay, we have to design a bunch of visualizations that seem somewhat realistic [... and not something that] you never really would see in the real world" (LG).

On top of the compromises we often have to make on the level of abstraction we can operate on while still staying somewhat realistic, we are faced with the the need to reduce noise when designing the questions in the test. For instance, as we pointed out in Sec. 4, visualization skills are difficult to isolate from other skills and knowledge.

> There will be a lot of noise coming from knowledge about this topic, knowledge about the dataset. Because when you ask a question about something, they don't necessarily have to use the chart. They can just get something out of their memory or something. So we have to be careful in that aspect. (BCK)

As a result, the design process for individual questions is often very iterative and involves qualitative work, which was evident in prior work in test development and evaluation [19, 29, 30, 41, 42, 49, 61]. Such careful iterations are necessary, because isolating the relevant visualization skill and reducing noise in the questions is crucial for the validity of a test. These challenges of test design can only increase as we shift to assessing higher-level visualization skills, such as visualization construction, visual decision-making, or visual metaphors, because the sources of noise inevitably increase as complexity increases (Sec. 4).

6.3 Calls to Action

Test designers face many challenges during the development process and often have to make compromises due to the complexity of the construct as mentioned in Sec. 4.1. The test modality can also constrain what the test is able to measure, thereby influencing the ecological validity of a test, as oftentimes what we *can* measure is not aligned with what we *need* to use visualizations for in the real world. We call on test designers to seek ecological validity, take an iterative test design approach over time, and lean on specialized knowledge to support test development and validation.

Improve Ecological Validity of Tests

When we create visualizations and questions for assessments, we must face difficult trade-offs between abstraction and realism. Translating tasks from theoretical taxonomies into concrete visual representations and instructions is not a trivial process and requires careful, iterative design processes. Our efforts to minimize noise from individuals' knowledge or contextual factors in tests may also come at the cost of reducing the ecological validity:

> I would like [...] higher-level and then also more realistic use cases, like the ways that people would use visualizations in their lives versus just kind of completely decontextualized. (MH)

> [In] health literacy, there's a decision making element, right? Because if you have a chart in that field, usually it's to make some diagnosis [... or] life changes to be healthier. And then versus something like Cleveland and McGill, there's not any decision really to be made. It's just making a judgment, but that's not teaching you how to apply that information in what you do day to day. So maybe that could mean having more realistic examples in these assessments. (KB)

To improve ecological validity, **we should situate measurements in real-world contexts**. Such measures could better reflect people's true abilities in visualizations.

Expand Assessment Modalities

Following our call to expand test scopes (Sec. 4.4), the format we use to assess these higher-level skills would have to change as well, as multiple choice questions are less likely to offer adequate support (Sec. 6.1). We should make an effort to **embrace alternative assessment formats**, such as using more open-ended questions. We expect the difficulty of isolating visualization skills to generally increase as the complexity of the ability we assess increases, due to the increasing sources of noise (e.g., complex assessment tools, technical skills). It is important, however, to not overly bind ourselves with the properties (e.g., logical, straightforwardness) that contributed to our progress in visualization literacy, as the nature of the construct itself (e.g., complex, fluid) often contradicts those properties. We must expand how we approach visualization literacy assessments, because "measurement is in a way influencing people's perceptions on the construct" (LG).

Seek Interdisciplinary Collaborations to Enhance Measurements

As an interdisciplinary field, visualization literacy can benefit a lot from inputs from related domains, such as education and psychology, to enrich our conceptualization of what visualization literacy is and to improve how we measure it. KB expressed:

> We need to work more closely with psychologists to actually get this empirical backing to all the stuff we do because they do this all the time [...] We can bring the CS side and the other things, and they can bring the actual psychology measurements. (KB)

Formal training in methods such as statistics and psychometrics testing that are necessary for analyzing and validating tests is still lacking in our community. **We should embrace opportunities to lean on specialized knowledge to support test development.** For example, psychologists have proposed theoretical and methodological work on ensuring cross-cultural validity [44, 48] when translating measurement instruments (something we are already attempting, as discussed in Sec. 5.3). Such work, however, might not be easily accessible to researchers in our field, as it requires specialized knowledge that is often outside our typical scope. **Cross-domain collaborations could address the many intricacies of psychometric measurements.** It is good that we are already making these connections with meetups, panels, and workshops, which we have already seen benefits from, such as this position paper. We could leverage these connections and continue to spur interdisciplinary conversations, which would be important for fostering collaboration across these domains.

7 Discussion

7.1 Calls for Mindset Changes in the Community

In this autoethnography we identified several *calls to action* from the conceptual, operational, and methodological perspectives. Here, we call for additional *mindset* and *attitude* changes that are more cross-cutting and applicable to the community at large.

Be ready to compromise between what you *need* to measure and you actually *can* measure, because the fluidity of visualization literacy makes it an ever-moving target. A single test that covers every dimension is difficult to achieve. AL recalled: "My initial study was [...] to touch different dimensions [...][in one study,] but things were messy, because the test was not so reliable." We may thus need to develop tests specific to a particular domain or context. However, too individualized an approach to test development may make it difficult for test users to get a holistic understanding of participants' visualization literacy. Regardless of the context or literacy component one focuses on, **authors should explicitly define which components of visualization literacy they study** to facilitate clear communication within the field and outside of it. At a certain point, we may need to simply acknowledge that there is a limit to what we can precisely measure.

Resist the urge to choose a measure just to conform to community standards. Due to the complexity of visualization literacy and the small set of existing assessments, it is easy to end up studying aspects of visualization literacy not covered by existing assessments. In such cases, **it is not appropriate to use a widely-adopted test just to check the "measure visualization literacy" box**. Visualization literacy tests should not be used mindlessly to conform to a perceived standard, just as user studies should not be mindlessly used to evaluate a visualization system if another evaluation is more appropriate [46]. While we acknowledge that it may be difficult to resist these normative expectations, particularly during peer review, we are hopeful that the community will be receptive to well-reasoned deviations from what might be considered standard.

Design and re-design tests on a longer time scale, iteratively. As we are far from solving the visualization literacy measurement problem, existing tests should not be viewed as "complete" products. **We should scrutinize existing tests**, especially when we use them in practice. Some areas for improvement can only be discovered through actual use, not via theorizing about how we should measure visualization literacy. Test users should report what worked and did not work whenever possible, which will inform test designers on areas for iterative improvement.

Take outward-facing responsibility for how tests are disseminated and maintained, because diverse user groups outside of academia are also using these tests (Sec. 5.1). Reflecting on the web version of VLAT and its role in connecting different audiences, BCK recalled:

> I still get, from time to time, emails [...] about VLAT, trying to correct some errors in the website, or just asking questions or just sending me a really good vibe about "Hey, thanks for making this". [...] That means people do need this. (BCK)

Potential test users who are outside of our field might not be as familiar with the nuances of visualization literacy as we are, so they are trusting us to develop valid assessments that they can adopt easily. One way of taking greater responsibility is to talk to users from other domains who are adopting our tests to learn how we can further improve how we measure visualization literacy.

7.2 A Reflection on Our Methodological Approach

Our approach to surface tacit thoughts relied on a combination of collaborative autoethnography, thematic analysis, and diffractive analysis. Although conducting autoethnography with a group adds rigor, it has known logistical challenges, including "establishing shared goals, agreeing how the group will work together, and meeting timelines" [58]. These issues could have been exacerbated in a paper with over a dozen authors across multiple continents on a tight timeline. We addressed these challenges through (1) a shared spreadsheet documenting assigned tasks and progress for all analysts, coordinated by the first author and visible to all—creating social accountability; (2) a group meeting where analysts practiced the analysis process on a short interview excerpt; and (3) a flow diagram analysts followed to keep procedures consistent.[8] These procedures helped us be transparent about progress and kept everyone on the same page about shared goals and timelines.

While we leaned on thematic analysis (Sec. 3.3) to formulate our tacit thoughts into a coherent narrative, we intentionally employed diffractive analysis to avoid passive agreement among ourselves. We explicitly added some distance between the analysts and the transcripts, which was our attempt at further reducing bias and increasing nuance to address the weakness of potentially lacking nuance in individual autoethnography [58]. Our approach to collaborative autoethnography created opportunities for a clash of perspectives and helped us distill the many factors that contribute to the wicked problem of visualization literacy measurement. However, we acknowledge that the majority of our perspectives stem from a Western-centric visualization philosophy. Future work could engage in more cross-cultural reflections, which would be especially valuable as visualization literacy itself is fluid across cultures.

8 Next Steps

In this autoethnographic reflection, we leaned on our professional and practical experiences as users and designers of visualization literacy tests to surface the tacit thoughts often left unreported in individual research papers. Despite our formulation of visualization literacy measurement as a wicked problem—one that is complex and without an exhaustive set of potential solutions [73]—we identified some concrete next steps to study visualization literacy more comprehensively. For example, test designers should expand the scope of their tests, both in terms of content and modality, and work towards improving the ecological validity of tests. Test users should report both what worked and what did not to help us collectively iterate on test design and how-to guides. We should develop centralized repositories for implementing tests. Most importantly, we should shift our mindsets to recognize the construct's fluidity and expect change in both our understanding of the construct and its measurements. As suggested by the formulation of this as a wicked problem, we do not claim to have an exhaustive set of rules or procedures for readers to follow. Instead, we advocate embracing the wickedness of this measurement problem, being willing to make compromises, and continually discussing—and updating—our conceptualizations of visualization literacy.

[8] Fig. 2 is a compact version of the diagram we used during analysis.

SUPPLEMENTAL MATERIALS

All supplemental materials are available on OSF at https://osf.io/xwr4c/, which include (1) interview protocol with the semi-structured questions, (2) share link to the Miro boards for analysis: https://miro.com/app/board/uXjVIfMDSTI=/?share_link_id=445344623741, (3) PDF version of the Miro boards.

ACKNOWLEDGMENTS

We extend our gratitude to all of our organizers and participants from the ACM CHI 2024 workshop on visualization literacy. This work was supported in part by the European Union – Next-Generation, Mission 4 Component 1 CUP D53D23008690006, project name: "Characterizing and Measuring Visual Information Literacy" ID 2022JJ3PA5, as well as NSF IIS-2403094. The following statements are included by Ge, in accordance with the NSF Graduate Research Fellowship Program Administrative Guide (NSF 25-033): This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-2234667. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

[1] P. Adelberger, O. Lesota, K. Eckelt, M. Schedl, and M. Streit. Iguanodon: A Code-Breaking Game for Improving Visualization Construction Literacy. *IEEE Trans. Visual Comput. Graphics*, 31(9):5713–5725, 2025. doi: 10.1109/TVCG.2024.3468948 1, 2, 4, 6

[2] D. Akbaba, D. Lange, M. Correll, A. Lex, and M. Meyer. Troubling Collaboration: Matters of Care for Visualization Design Study. In *Proc. CHI*, CHI '23, article no. 812, 15 pages. ACM, New York, 2023. doi: 10.1145/3544548.3581168 2

[3] J. Alford and B. W. Head. Wicked and less wicked problems: a typology and a contingency framework. *Policy and Society*, 36(3):397–413, 2017. doi: 10.1080/14494035.2017.1361634 2

[4] B. Alper, N. H. Riche, F. Chevalier, J. Boy, and M. Sezgin. Visualization Literacy at Elementary School. In *Proc. CHI*, CHI '17, 13 pages, p. 5485–5497. ACM, New York, 2017. doi: 10.1145/3025453.3025877 1, 2, 6

[5] R. Amar, J. Eagan, and J. Stasko. Low-Level Components of Analytic Activity in Information Visualization. In *Proceedings of the Proceedings of the 2005 IEEE Symposium on Information Visualization*, INFOVIS '05, p. 15. IEEE Computer Society, USA, 2005. doi: 10.1109/INFOVIS.2005.24 4

[6] H. S. Andrianarivony, T. D. Raharison, I. T. Rakotoniaina, M. T. N. Ramarokoto, N. Rakotondravony, and L. Harrison. Investigating the Use of Native and Secondary Language with Data Visualization in Madagascar. In *IEEE VIS Posters*. Oklahoma City, 2022. Poster paper. 3

[7] B. Bach, M. Keck, F. Rajabiyazdi, T. Losev, I. Meirelles, J. Dykes, R. S. Laramee, M. AlKadi, C. Stoiber, S. Huron, C. Perin, L. Morais, W. Aigner, D. Kosminsky, M. Boucher, S. Knudsen, A. Manataki, J. Aerts, U. Hinrichs, J. C. Roberts, and S. Carpendale. Challenges and Opportunities in Data Visualization Education: A Call to Action. *IEEE Trans. Visual Comput. Graphics*, 30(1):649–660, 2024. doi: 10.1109/TVCG.2023.3327378 2, 4

[8] K. Barad. *Meeting the Universe Halfway: Quantum Physics and the Entanglement of Matter and Meaning*. Duke University Press, 2006. doi: 10.1215/9780822388128 2, 3

[9] K. Barad. Diffracting Diffraction: Cutting Together-Apart. *Parallax*, 20(3):168–187, 2014. doi: 10.1080/13534645.2014.927623 2

[10] S. Beschi, D. Falessi, S. Golia, and A. Locoro. Characterizing Data Visualization Literacy for Standardization: A Systematic Literature Review. *IEEE Access*, 13:65704–65725, 2025. doi: 10.1109/ACCESS.2025.3559298 3

[11] S. Beschi, S. Golia, D. Falessi, and A. Locoro. Envisioning Visual Information Literacy: some reflection on conceptual and practical stances. https://osf.io/kbaes, 2024. 3

[12] R. Birchfield, M. Caten, E. Cheng, M. Kelly, T. Larson, H. P. Pham, Y. Ding, N. Rakotondravony, and L. Harrison. VisQuiz: Exploring Feedback Mechanisms to Improve Graphical Perception. In *2022 IEEE Visualization and Visual Analytics (VIS)*, pp. 95–99, 2022. doi: 10.1109/VIS54862.2022.00028 1, 3

[13] F. Bishop, J. Zagermann, U. Pfeil, G. Sanderson, H. Reiterer, and U. Hinrichs. Construct-A-Vis: Exploring the Free-Form Visualization Processes of Children. *IEEE Trans. Visual Comput. Graphics*, 26(1):451–460, 2020. doi: 10.1109/TVCG.2019.2934804 2, 6

[14] J. Boy, R. A. Rensink, E. Bertini, and J.-D. Fekete. A Principled Way of Assessing Visualization Literacy. *IEEE Trans Vis Comput Graph*, 20(12):1963–1972, 2014. doi: 10.1109/TVCG.2014.2346984 1, 2, 4, 7

[15] M. Brehmer and T. Munzner. A Multi-Level Typology of Abstract Visualization Tasks. *IEEE Trans. Visual Comput. Graphics*, 19(12):2376–2385, 2013. doi: 10.1109/TVCG.2013.124 4

[16] K. Börner, A. Bueckle, and M. Ginda. Data visualization literacy: Definitions, conceptual frameworks, exercises, and assessments. *Proceedings of the National Academy of Sciences*, 116(6):1857–1864, 2019. doi: 10.1073/pnas.1807180116 1, 2

[17] K. Börner, A. Maltese, R. N. Balliet, and J. Heimlich. Investigating aspects of data visualization literacy using 20 information visualizations and 273 science museum visitors. *Information Visualization*, 15(3):198–213, 2016. doi: 10.1177/1473871615594652 1, 2

[18] A.-F. Cabouat, T. He, F. Cabric, T. Isenberg, and P. Isenberg. Position paper: A case to study the relationship between data visualization readability and visualization literacy. https://hal.science/hal-04523790, 2024. 3

[19] A.-F. Cabouat, T. He, P. Isenberg, and T. Isenberg. PREVis: Perceived Readability Evaluation for Visualizations. *IEEE Trans. Visual Comput. Graphics*, 31(1):1083–1093, 2025. doi: 10.1109/TVCG.2024.3456318 3, 8

[20] R. A. Carr-Hill and J. Pessoa. *International Literacy Statistics: A Review of Concepts, Methodology and Current Data*. UNESCO Institute for Statistics Montreal, 2008. 2

[21] C. Cazden, B. Cope, N. Fairclough, J. Gee, M. Kalantzis, G. Kress, A. Luke, C. Luke, S. Michaels, and M. Nakata. A Pedagogy of Multiliteracies: Designing Social Futures. *Harvard educational review*, 66(1):60–92, 1996. 2, 4

[22] H. Chang, F. Ngunjiri, and K.-A. C. Hernandez. *Collaborative Autoethnography*. Routledge, New York, 2016. doi: 10.4324/9781315432137 2

[23] V. Clarke and V. Braun. Thematic analysis. *The Journal of Positive Psychology*, 12(3):297–298, 2017. doi: 10.1080/17439760.2016.1262613 3

[24] R. J. Cohen, W. J. Schneider, and R. Tobin. *Psychological Testing and Assessment*. McGraw Hill LLC, New York, NY, 10th ed., 2022. 7

[25] B. Cope and M. Kalantzis. *The Things You Do to Know: An Introduction to the Pedagogy of Multiliteracies*, pp. 1–36. Palgrave Macmillan UK, London, 2015. doi: 10.1057/9781137539724_1 2, 4

[26] M. Correll. The Mythos of Visualization Literacy. https://mcorrell.medium.com/the-mythos-of-visualization-literacy-5dd02db56254, 2018. 3

[27] M. Correll. What Does "Visualization Literacy"' Mean, Anyway? https://medium.com/multiple-views-visualization-research-explained/what-does-visualization-literacy-mean-anyway-22f3b3badc0, 2018. 3, 5

[28] M. Creamer, L. Padilla, and M. Borkin. Finding Gaps in Modern Visualization Literacy, 2024. doi: 10.31219/osf.io/jy9v2 3

[29] Y. Cui, L. W. Ge, Y. Ding, L. Harrison, F. Yang, and M. Kay. Promises and Pitfalls: Using Large Language Models to Generate Visualization Items. *IEEE Trans. Visual Comput. Graphics*, 31(1):1094–1104, 2025. doi: 10.1109/TVCG.2024.3456309 3, 8

[30] Y. Cui, L. W. Ge, Y. Ding, F. Yang, L. Harrison, and M. Kay. Adaptive Assessment of Visualization Literacy. *IEEE Trans. Visual Comput. Graphics*, 30(1):628–637, 2024. doi: 10.1109/TVCG.2023.3327165 2, 3, 6, 8

[31] F. Daneshzand, C. Perin, and S. Carpendale. Design Process of 'Shredded Lives': An Illustrated Exploration. In *2024 IEEE VIS Arts Program (VISAP)*, pp. 39–52, 2024. doi: 10.1109/VISAP64569.2024.00008 2

[32] R. Davis, X. Pu, Y. Ding, B. D. Hall, K. Bonilla, M. Feng, M. Kay, and L. Harrison. The Risks of Ranking: Revisiting Graphical Perception to Model Individual Differences in Visualization Performance. *IEEE Trans. Visual Comput. Graphics*, 30(3):1756–1771, 2024. doi: 10.1109/TVCG.2022.3226463 3

[33] Y. Ding, J. Wilburn, H. Shrestha, A. Ndlovu, K. Gadhave, C. Nobre, A. Lex, and L. Harrison. reVISit: Supporting Scalable Evaluation of Interactive Visualizations. In *2023 IEEE Visualization and Visual Analytics (VIS)*, pp. 31–35, 2023. doi: 10.1109/VIS54172.2023.00015 7

[34] C. Ellis. *The Ethnographic I: A Methodological Novel about Autoethnography*. AltaMira Press, 2004. 2

[35] C. Ellis, T. E. Adams, and A. P. Bochner. Autoethnography: An Overview. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*,

12(1), 2010. doi: 10.17169/fqs-12.1.1589 2

[36] E. E. Firat, A. Denisova, and R. S. Laramee. Treemap Literacy: A Classroom-Based Investigation. In M. Romero and B. Sousa Santos, eds., *Eurographics 2020 - Education Papers*. The Eurographics Association, 2020. doi: 10.2312/eged.20201032 2

[37] E. E. Firat, A. Denisova, M. L. Wilson, and R. S. Laramee. P-Lite: A study of parallel coordinate plot literacy. *Visual Informatics*, 6(3):81–99, 2022. doi: 10.1016/j.visinf.2022.05.002 2

[38] E. E. Firat, A. Joshi, and R. S. Laramee. Interactive visualization literacy: The state-of-the-art. *Information Visualization*, 21(3):285–310, 2022. doi: 10.1177/14738716221081831 2

[39] J. Gäbler, C. Winkler, N. Lengyel, W. Aigner, C. Stoiber, G. Wallner, and S. Kriglstein. Diagram Safari: A Visualization Literacy Game for Young Children. In *Proc. CHI PLAY EA*, CHI PLAY '19 Extended Abstracts, 8 pages, p. 389–396. ACM, New York, 2019. doi: 10.1145/3341215.3356283 2, 6

[40] M. Galesic and R. Garcia-Retamero. Graph Literacy: A Cross-Cultural Comparison. *Medical Decision Making*, 31(3):444–457, 2011. doi: 10.1177/0272989X10373805 1, 2, 6

[41] L. W. Ge, Y. Cui, and M. Kay. CALVI: Critical Thinking Assessment for Literacy in Visualizations. In *Proc. CHI*, CHI '23, article no. 815, 18 pages. ACM, New York, 2023. doi: 10.1145/3544548.3581406 1, 2, 3, 6, 8

[42] L. W. Ge, Y. Cui, and M. Kay. AVEC: An Assessment of Visual Encoding Ability in Visualization Construction. In *Proc. CHI*, CHI '25, article no. 1166, 16 pages. ACM, New York, 2025. doi: 10.1145/3706598.3713364 1, 2, 3, 4, 6, 8

[43] L. W. Ge, M. Hedayati, Y. Cui, Y. Ding, K. Bonilla, A. Joshi, A. Ottley, B. Bach, B. C. Kwon, D. N. Rapp, E. Peck, L. M. Padilla, M. Correll, M. A. Borkin, L. Harrison, and M. Kay. Toward a More Comprehensive Understanding of Visualization Literacy. In *CHI EA*, CHI EA '24, article no. 494, 7 pages. ACM, New York, 2024. doi: 10.1145/3613905.3636289 2, 3

[44] K. F. Geisinger. Cross-cultural normative assessment: Translation and adaptation issues influencing the normative interpretation of assessment instruments. *Psychological Assessment*, 6(4):304–312, 1994. doi: 10.1037/1040-3590.6.4.304 9

[45] P. Geist-Martin, L. Gates, L. Wiering, E. Kirby, R. Houston, A. Lilly, and J. Moreno. Exemplifying Collaborative Autoethnographic Practice via Shared Stories of Mothering. *Journal of Research Practice*, 6(1):M8–M8, 2010. 2

[46] S. Greenberg and B. Buxton. Usability evaluation considered harmful (some of the time). In *Proc. CHI*, CHI '08, 10 pages, p. 111–120. ACM, New York, 2008. doi: 10.1145/1357054.1357074 9

[47] A. Hakone, L. Harrison, A. Ottley, N. Winters, C. Gutheil, P. K. J. Han, and R. Chang. PROACT: Iterative Design of a Patient-Centered Visualization for Effective Prostate Cancer Health Risk Communication. *IEEE Trans. Visual Comput. Graphics*, 23(1):601–610, 2017. doi: 10.1109/TVCG.2016.2598588 3

[48] R. K. Hambleton and A. Kanjee. Increasing the Validity of Cross-Cultural Assessments: Use of Improved Methods for Test Adaptations. *European Journal of Psychological Assessment*, 11(3):147–157, 1995. doi: 10.1027/1015-5759.11.3.147 9

[49] T. He, P. Isenberg, R. Dachselt, and T. Isenberg. BeauVis: A Validated Scale for Measuring the Aesthetic Pleasure of Visual Representations. *IEEE Trans. Visual Comput. Graphics*, 29(1):363–373, 2023. doi: 10.1109/TVCG.2022.3209390 3, 8

[50] M. Hedayati, A. Hunt, and M. Kay. From pixels to practices: Reconceptualizing visualization literacy, 2024. doi: 10.31219/osf.io/6mq42 1, 2, 3

[51] M. Hedayati and M. Kay. What University Students Learn In Visualization Classes. *IEEE Trans. Visual Comput. Graphics*, 31(1):1072–1082, 2025. doi: 10.1109/TVCG.2024.3456291 3

[52] E. Huynh, A. Nyhout, P. Ganea, and F. Chevalier. Designing Narrative-Focused Role-Playing Games for Visualization Literacy in Young Children. *IEEE Trans. Visual Comput. Graphics*, 27(2):924–934, 2021. doi: 10.1109/TVCG.2020.3030464 1

[53] S. Jones. *The Sage Handbook of Qualitative Research*, chap. Autoethnography: Making the personal political. Sage, 2005. 2

[54] A. Joshi, K. Börner, R. S. Laramee, L. Harrison, E. E. Firat, and B. C. Kwon. Visualization literacy for general audiences-can we make a difference? IEEE VIS 2021 Panel. 2, 3

[55] M. Keck, S. Huron, G. Panagiotidou, C. Stoiber, F. Rajabiyazdi, C. Perin, J. C. Roberts, and B. Bach. EduVis: Workshop on Visualization Education, Literacy, and Activities. `https://arxiv.org/abs/2303.10708`, 2023. 2

[56] B. C. Kwon and B. Lee. A Comparative Evaluation on Online Learning Approaches Using Parallel Coordinate Visualization. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, 5 pages, p. 993–997. ACM, New York, 2016. doi: 10.1145/2858036.2858101 1, 3

[57] K. Lai, J. Cabrera, J. M. Vitale, J. Madhok, R. Tinker, and M. C. Linn. Measuring Graph Comprehension, Critique, and Construction in Science. *Journal of Science Education and Technology*, 25:665–681, 2016. doi: 10.1007/s10956-016-9621-9 2

[58] J. C. Lapadat. Ethics in Autoethnography and Collaborative Autoethnography. *Qualitative Inquiry*, 23(8):589–603, 2017. doi: 10.1177/1077800417704462 2, 9

[59] J. C. Lapadat, N. E. Black, P. G. Clark, R. M. Gremm, L. W. Karanja, L. W. Mieke, and L. Quinlan. Life Challenge Memory Work: Using Collaborative Autobiography to Understand Ourselves. *International Journal of Qualitative Methods*, 9(1):77–104, 2010. doi: 10.1177/160940691000900108 2

[60] A. Lazar, B. Jelen, A. Pradhan, and K. A. Siek. Adopting Diffractive Reading to Advance HCI Research: A Case Study on Technology for Aging. *ACM Trans. Comput.-Hum. Interact.*, 28(5), article no. 32, 29 pages, 2021. doi: 10.1145/3462326 2

[61] S. Lee, S.-H. Kim, and B. C. Kwon. VLAT: Development of a Visualization Literacy Assessment Test. *IEEE Trans. Visual Comput. Graphics*, 23(1):551–560, 2017. doi: 10.1109/TVCG.2016.2598920 1, 2, 3, 7, 8

[62] S. Lee, B. C. Kwon, J. Yang, B. C. Lee, and S.-H. Kim. The Correlation between Users' Cognitive Characteristics and Visualization Literacy. *Applied Sciences*, 9(3), 2019. doi: 10.3390/app9030488 3

[63] Z. Liu, R. J. Crouser, and A. Ottley. Survey on Individual Differences in Visualization. *Computer Graphics Forum*, 39(3):693–712, 2020. doi: 10.1111/cgf.14033 4

[64] D. Lloyd and J. Dykes. Human-Centered Approaches in Geovisualization Design: Investigating Multiple Methods Through a Long-Term Case Study. *IEEE Trans. Visual Comput. Graphics*, 17(12):2498–2507, 2011. doi: 10.1109/TVCG.2011.209 2

[65] A. Locoro, W. P. Fisher, and L. Mari. Visual Information Literacy: Definition, Construct Modeling and Assessment. *IEEE Access*, 9:71053–71071, 2021. doi: 10.1109/ACCESS.2021.3078429 3

[66] H. Mansoor and L. Harrison. *Data Visualization Literacy and Visualization Biases: Cases for Merging Parallel Threads*, pp. 87–96. Springer International Publishing, Cham, 2018. doi: 10.1007/978-3-319-95831-6_7 3

[67] M. L. McHugh. Interrater reliability: the kappa statistic. *Biochemia medica*, 22(3):276–282, 2012. 3

[68] Y. Okan, E. Janssen, M. Galesic, and E. A. Waters. Using the Short Graph Literacy Scale to Predict Precursors of Health Behavior Change. *Medical Decision Making*, 39(3):183–195, 2019. doi: 10.1177/0272989X19829728 2

[69] J. Otto. Visualization Notebooks are Paper Tools. `https://osf.io/z2u3w`, 2024. 3

[70] J. T. Otto and S. Davidoff. Visualization Artifacts are Boundary Objects. In *ProcḂELIV*, pp. 81–88, 2024. doi: 10.1109/BELIV64461.2024.00014 3

[71] S. Pandey and A. Ottley. Mini-VLAT: A Short and Effective Measure of Visualization Literacy. *Computer Graphics Forum*, 42(3):1–11, 2023. doi: 10.1111/cgf.14809 2, 6

[72] F. Rajabiyazdi, M. Keck, C. Stoiber, J. Roberts, H. Subramonyam, L. Ge, M. Boucher, B. Bach, L. Besancon, M. Brossier, et al. EduVis: 2nd IEEE VIS Workshop on Visualization Education, Literacy, and Activities. `https://ieee-eduvis.github.io/IEEE-EduVis-Workshop-2024.pdf`. 2

[73] H. W. Rittel and M. M. Webber. Dilemmas in a general theory of planning. *Policy sciences*, 4(2):155–169, 1973. doi: 10.1007/BF01405730 2, 9

[74] M. Solen. Scoping the Future of Visualization Literacy: A Review. `https://osf.io/eypgm`, 2022. 1, 2