Mini-VLAT: A Short and Effective Measure of Visualization Literacy

Saugat Pandey and Alvitta Ottley

Washington University in St. Louis, St. Louis, MO, USA

Abstract

The visualization community regards visualization literacy as a necessary skill. Yet, despite the recent increase in research into visualization literacy by the education and visualization communities, we lack practical and time-effective instruments for the widespread measurements of people's comprehension and interpretation of visual designs. We present Mini-VLAT, a brief but practical visualization literacy test. The Mini-VLAT is a 12-item short form of the 53-item Visualization Literacy Assessment Test (VLAT). The Mini-VLAT is reliable (coefficient omega = 0.72) and strongly correlates with the VLAT. Five visualization experts validated the Mini-VLAT items, yielding an average content validity ratio (CVR) of 0.6. We further validate Mini-VLAT by demonstrating a strong positive correlation between study participants' Mini-VLAT scores and their aptitude for learning an unfamiliar visualization using a Parallel Coordinate Plot test. Overall, the Mini-VLAT items showed a similar pattern of validity and reliability as the 53-item VLAT. The results show that Mini-VLAT is a psychometrically sound and practical short measure of visualization literacy.

CCS Concepts

• *Human-centered computing* → *Visualization tools, Empirical Study;*

1. INTRODUCTION

The proliferation of visualization technologies, toolkits, journalism, and social media has led to the pervasiveness of visualization in people's daily lives. However, people's capacity to interpret and use graphics varies [PYH*12, Ott20, LCO20], leaving the visualization community with unsolved questions with far-reaching consequences. With this increased use of visualization tools to convey complex data, measuring visualization literacy (VL) is becoming increasingly vital.

VL is a crucial skill for understanding and interpreting complex imagery communicated by interactive visual designs [BHH*21]. As a result, researchers in the Information Visualization community have attempted to examine users' visualization literacy [JBL*21, BBG19, BRBF14, PKH21, Car92]. Across several studies, scholars have proposed methods to assess visualization literacy, attempted to understand the current state of data visualization comprehension among the general public [PAEE19, BBG19], and suggested various ways of learning unfamiliar visualizations to improve users' visualization literacy.

Although VL has been described as an essential skill through various workshops, keynote talks, and research works [BHH*21, JBL*21], researchers still lack a validated and accurate tool for testing user visualization literacy that can be widely deployed [Yi12]. Perhaps the most well-known publicly available visualization literacy test, Visualization Literacy Assessment Test (VLAT) [LKK16],

consists of 53 questions across 12 different visualization types. Each question has a limit of 25 seconds, making it approximately 22 minutes long. Such a long test typically necessitates strong cognitive demands, which might lead to cognitive tiredness and erroneous answers for some respondents [Sit22]. Thus, the time commitment involved with VLAT limits its use to primarily academic settings.

The recent discussions about visualization literacy at IEEE VIS 2022 focused on developing methods to test visualization literacy and an appropriate alternative for VLAT. One of the issues raised was the need for a short and effective visualization literacy test that can be used for large-scale data collection and visualization evaluation. *To address this need, we developed an abbreviated version of the Visualization Literacy Assessment Test, Mini-VLAT.* The Mini-VLAT consists of 12 multiple-choice questions spread across 12 different visualization types, making it 5 minutes long. We evaluated the Mini-VLAT with the established procedure of test development in Psychology and Established Measurement [CSP96] and other highly cited works in the field of psychology [Sit22, SMA00] when creating a shortened test in 5 phases: (1) Test Item Generation (2) Piloting and Item Refinement (3) Reliability Evaluation (4) Correlation between Mini-VLAT and VLAT. (5) Testing Mini-VLAT's predictive capacity.

In summary, this paper presents Mini-VLAT, a brief and practical test for measuring visualization literacy that can be widely

used. We follow established procedures for test development and validation, including test generation, reliability evaluation, correlation analysis with the established VLAT, and testing the predictive capacity of Mini-VLAT. Our results demonstrate that Mini-VLAT is a psychometrically sound and reliable measure of visualization literacy that is strongly correlated with VLAT and has similar validity and reliability properties. The online version of Mini-VLAT can be found at https://washuvis.github.io/minivlat/.

2. RELATED WORK

Substantial work has been done around defining and examining visualization literacy. In this section, we discuss previous works on defining visualization literacy and then discuss relevant works that examined visualization literacy.

2.1. Definitions of Visualization Literacy

Several seminars at visualization conferences have sparked conversations among scholars to define visualization literacy. It intersects with several disciplines of study, including cognitive psychology and education. Although several researchers have attempted to define visualization literacy, there is no clear consensus.

In an early study, graph literacy was used instead of visualization literacy [GGR11]. Galesic et al. [GGR11] created a test scale to assess graph literacy in the context of health. The test items were divided into three comprehension levels: reading the data, reading between the data, and reading beyond the data. However, Boy et al. [BRBF14] laid the groundwork for visualization literacy in the Information Visualization community. They defined visualization literacy as

> "The ability to confidently use a given data visualization to translate questions specified in the data domain into visual queries in the visual domain, as well as interpreting visual patterns in the visual domain as properties in the data domain." [BRBF14]

In 2015, Börner et al. [BMBH16] investigated the familiarity of different visualization types among youth and adult museum visitors. They defined visualization literacy as

> "The ability to make meaning from and interpret patterns, trends, and correlations in visual representations of data." [BMBH16]

Based on these definitions, Lee et al. [LKK16] devised their own definition of visualization literacy and introduced *Visualization Literacy Assessment Test (VLAT)*. They defined visualization literacy as

> "the ability and skills to read and interpret visually represented data in and to extract information from data visualizations." [LKK16]

We assert that visualization literacy is a multidimensional construct, and measuring its full scope with a single scale is untenable. Thus, this work focuses on *measuring one's ability to read and interpret visually represented data*, which we argue is a necessary skill for every proposed definition of VL.

2.2. Measuring Visualization Literacy

Several research works have been proposed towards understanding and measuring visualization literacy [GGR11, FDWL22, FDL20, LKK16, KBC01, BMBH16]. Most early works have measured the ability to read and interpret a graph [GGR11, Ber10, Car92, Wai92]. However, the question items in these works consisted of primitive graphs based on the three comprehension levels mentioned before.

Differing from this prior work, Boy et al. [BRBF14] applied item response theory to measure visualization literacy. The items in the test were based on six visualization tasks (minimum, maximum, variation, intersection, average, and comparison). The test included line charts, bar charts, and scatterplots.

Börner et al. [BMBH16] applied yet another approach to measuring data visualization literacy by examining the familiarity of various visualization designs. Three US science museums hosted this investigation. Börner et al. [BMBH16] selected 20 visualizations from textbooks and popular web visualization libraries, such as the D3.js library. The selected visualizations contain two charts, five maps, eight graphs, and five network layouts. Five of the twenty visual designs were shown to visitors, who were asked to express their knowledge of the visual designs and identify the design's name.

Although this work does not attempt to create a visualization literacy measure, it does provide insight into how people had significant limitations in identifying and understanding different data visualizations. Subsequently, Börner et al. [BBG19] published a visualization literacy framework paper that included a set of definitions, conceptual frameworks, exercises, and assessments to describe visualization literacy. They introduced a data visualization literacy framework (DVL-FM) based on seven core hierarchical typologies of extracting insights from data [BBG19].

In 2016, Lee et al. [LKK16] developed a visualization literacy assessment test (VLAT) that consists of 12 data visualizations and 53 multiple-choice test items. This test demonstrated high reliability and validity. The test was also validated among five visualization experts. The VLAT encouraged researchers to extend the work by adding more advanced visual designs [FDL20, FDWL22]. Firat et al. [FDWL22] developed a parallel coordinates literacy test, P-Lite, with diverse images generated using popular PCP software tools. Like VLAT, P-Lite was created to test the user's comprehension and ability to interpret high-dimensional visual designs like parallel coordinates. Firat et al. [FDL20] also created a treemap literacy test to assess the user's understanding and ability to interpret treemaps.

However, the length of the VLAT serves as its limitation. It limits the researchers in the visualization community to investigate the users' ability to read and interpret visually represented data on a larger scale. In this study, we presented an abbreviated version of the large VLAT that has comparable reliability and validity as the VLAT with 12 items.

3. WHY 12-ITEM Mini-VLAT?

Underscoring the need for a shorter visualization literacy test, some works have attempted to shorten the VLAT for research purposes [MVH17, LKY*19]. However, none of them followed the

guidelines concerning the development of short forms [SMA00, SSBS02]. For example, Lee et al. [LKY*19] reduced the number of items from 53 to 41 to demonstrate the correlation between the users' cognitive characteristics and visualization literacy. They shortened that test by removing the 12 items with the lowest discriminating values, i.e., the items that do not reliably distinguish the high and low performers on the test. Mansoor et al. [MVH17] also decreased the number of items from 53 to 22, citing the participants' performance degradation over time due to fatigue in their pilot study. However, a 22-item or 41-item test could still be too long for some studies.

What distinguishes VLAT from other visualization literacy assessments is how comprehensive it is by testing 12 different visualization designs with up to 7 types of questions for each. Inevitably, a short form will produce a less complete test, and we weighed the tradeoff between maintaining the variety of visualization designs or the task types. To create an abbreviated test that can be used in a wide range of non-experts settings, we opted to maintain the diversity of the visualization types. We hypothesize that we could produce a reliable and valid abbreviated test by selecting only one item from each visualization type if done carefully. It will also significantly decrease the completion time, making it suitable for widespread use when measuring visualization literacy. In the following sections, we describe the Mini-VLAT development process in detail.

4. DEVELOPMENT OF THE Mini-VLAT

The literature review demonstrates several existing methods for examining people's ability to read, interpret, and extract information from data visualization. However, we assert that a short but valid measure will be more practical for widespread use. Ultimately, we aim to develop a tool for researchers and visualization designers to gauge whether people can read visualization and track how visualization literacy changes over time. To achieve this, we opted to shorten an existing survey and leverage best practices in psychology to develop short-form surveys [SMA00, Sit22, DOBL06, Mel87].

We accomplish this through a 5-phase process:

Phase I: We began by replicating the VLAT and computing the total-item correlation for each item in the test. We then invited five domain experts in the Information Visualization community to evaluate the items for the large Mini-VLAT [Ass74, MH06]. We selected the item with high item-total correlation and positive *CVR* from each visualization type for the Mini-VLAT.

Phase II: The items selected in Phase I were tried on a sample of test takers. Based on the results from the analysis, we made revisions to the questions and included them in the final set of the Mini-VLAT.

Phase III: The items in the Mini-VLAT were evaluated in terms of reliability. We used the most commonly used internal consistency reliability measure called *reliability coefficient omega (ω)*. The reliability evaluation's findings provide additional validation for the validity that visualization literacy is measured accurately and consistently [Ass74].

Phase IV: It is critical to demonstrate sufficient variation between the short and full forms. Research studies generally calculate the correlation between the short and long forms based on a single test administration [CW98, LD96]. Smith et al. [SMA00] called this a methodological error because it leads to an overestimation of the correlation between the two forms. The answers to every question on the short form are counted twice, appearing on both sides of the correlation. So, naturally, any error or random variation in the responses to any of the short-form items is completely replicated in the long–form. By using this method, one is in fact correlating error to itself to some extent. In this phase, we compared the scores in the large Mini-VLAT and VLAT using Pearson's product-moment correlation coefficient. We established a strong positive correlation between the two forms.

Phase V: For our final validation test, we test how well the scores on the Mini-VLAT predict the performance on an independent aptitude test. For the aptitude test, we used the items from the Parallel Coordinate Test, P-Lite [FDWL22].

In the following sections, we describe the five stages of construction and evaluate the large Mini-VLAT.

4.1. Phase I: Test Item Selection

In this phase, we selected the items for the Mini-VLAT using expert feedback and the item-total correlation. Before choosing the items, we conducted a pilot study to replicate the Mini-VLAT and compute the item-total correlation for each item.

4.1.1. Replicating VLAT

The responses from the original study are not publicly available. As a result, we replicated the VLAT using the images and questions published at https://vlat.herokuapp.com/. After the replication, we wanted to ensure that the responses obtained in the replicated version significantly overlapped with the original study. Thus, we reached out to the authors of the VLAT, and they shared the data from their pilot study conducted among 191 participants. The horizontal error bars in Table 1 show that the responses obtained from the replication study for most of the items overlapped with the original ones. It shows that the questions in the VLAT are strongly reliable, producing stable and consistent results. Therefore, we decided to use the data from the replicated study to select the items for the Mini-VLAT. The pre-registration for this replication study and the data are available at https://osf.io/dy67k/.

Participants A total of 200 participants were recruited through Prolific for the replication. We applied the same requirements as the participants in VLAT [LKK16]. The participants had a HIT approval rate of 95% or higher and were limited to the United States. We only accepted participants who were native English speakers. As stated in the original study, we allotted 25 seconds to each question. Out of 200, we only ruled out one participant who was a random clicker. Eventually, a total of 199 participants remained. They comprised 112 females and 85 males with an age

Table 1: *The test items in the VLAT and the item-total correlation (r) and content validity ratio (CVR) of each item. An item with highest item-total correlation for each visualization type. One item from each visualization type initially selected for the Mini-VLAT. Comparison between the responses from the original and replicated study for each item using error bars (95% CI).*

Item ID	Visualization Type	Question	r	CVR	Original vs. Replicated VLAT
1		What was the price of a barrel of oil in February 2015?	0.43	0.2	
2		In which month was the price of a barrel of oil the lowest in 2015?	0.15	0.2	
3	Line Chart	What was the price range of a barrel of oil in 2015?	0.59	-0.2	
4		Over the course of the second half of 2015, the price of a barrel of oil was _____.	0.22	-0.2	
5		About how much did the price of a barrel of oil fall from April to September in 2015?	0.36	0.2	
6		What is the average internet speed in Japan?	0.29	0.6	
7	Bar Chart	In which country is the average internet speed the fastest in Asia?	0.12	0.6	
8		What is the range of the average internet speed in Asia?	0.47	-0.6	
9		How many countries in Asia is the average internet speed slower than Thailand?	0.03	-0.2	
10		What is the cost of peanuts in Las Vegas?	0.33	0.2	
11		About what is the ratio of the cost of a sandwich to the total cost of room service in Seattle?	0.26	0.6	
12	Stacked Bar Chart	In which city is the cost of soda the highest?	0.37	-0.6	
13		The cost of vodka in Atlanta is higher than that of Honolulu.	0.27	0.6	
14		The ratio of the cost of Soda to the cost of Water in Orlando is higher than that of Washington D.C.	0.23	-0.2	
15		What is the approval rating of Republicans among the people who have the education level of Postgraduate Study?	0.35	0.2	
16	100% Stacked Bar Chart	What is the education level of people in which the Democrats have the lowest approval rating?	0.37	1	
17		The approval rating of Republicans for the people who have the education level of Some College Degree is lower than that for the people who have the education level of Postgraduate Study.	0.20	0.6	
18		About what is the global smartphone market share of Samsung?	0.47	1	
19	Pie Chart	In which company is the global smartphone market share the smallest?	0.08	1	
20		The global smartphone market share of Apple is larger than that of Huawei.	0.13	0.2	
21		How many people have rated the taxi between 4.0 and 4.2?	0.23	0.6	
22	Histogram	What is the rating that the people have rated the taxi the most?	0.27	1	
23		More people have rated the taxi between 4.6 and 4.8 than between 4.2 and 4.4.	0.23	-0.2	
24		What is the weight for the person who is 165.1 cm tall?	0.41	0.2	
25		What is the height for the tallest person among the 85 males?	0.51	-0.2	
26		What is the range in weight for the 85 males?	0.36	-0.2	
27	Scatterplot	What is the height for a person who lies outside the others the most?	0.13	0.2	
28		A group of males are gathered around the height of 176 cm and the weight of 70 kg.	0.33	0.6	
29		There is a negative linear relationship between the height and the weight of the 85 males	0.43	0.6	
30		The weights for males with the height of 188 cm are all the same.	0.19	-0.2	
31		What was the average price of a pound of coffee beans in September 2013?	0.32	0.2	
32		When was the average price of a pound of coffee beans at minimum?	0.27	0.2	
33	Area Chart	What was the range of the average price of a pound of coffee beans between January 2013 and December 2014?	0.17	-0.6	
34		Over the course of 2013, the average price of a pound of coffee beans was _____.	0.27	-1	
35		What was the number of girls named 'Amelia' in 2010 in the UK?	0.22	-0.2	
36		About what was the ratio of the number of girls named 'Olivia' to those named 'Isla' in 2014 in the UK?	0.38	0.6	
37		Over the course of years between 2009 and 2014, when was the number of girls named 'Amelia' at the maximum?	0.29	0.6	
38	Stacked Area Chart	The number of girls named 'Isla' was _____ from 2009 to 2012.	0.18	-1	
39		In the UK, the number of girls named 'Amelia' in 2014 was more than it was in 2013.	0.06	-0.6	
40		Over the course of years between 2009 and 2014, the number of girls named 'Isla' was always more than 'Olivia'.	0.23	0.2	
41		What is the total length of the metro system in Beijing?	0.15	-0.2	
42		Which city's metro system has the largest number of stations?	0.52	0.2	
43		What is the range of the total length of the metro systems?	0.53	-0.6	
44		Which city's metro system does lie outside the relationship between the total system length and the number of stations most?	0.38	0.2	
45	Bubble Chart	A group of the metro systems of the world has approximately 300 stations and around a 200 km system length.	0.39	0.2	
46		In general, the ridership of the metro system increases as the number of stations increases.	0.05	1	
47		The metro system in Shanghai has more ridership than the metro system in Beijing.	0.48	0.2	
48		What was the unemployment rate for Indiana (IN) in 2015?	0.24	0.2	
49	Choropleth Map	In which state was the unemployment rate the highest in 2015?	0.14	0.2	
50		In 2015, the unemployment rate for Washington (WA) was higher than that of Wisconsin (WI).	0.33	0.6	
51		For which website was the number of unique visitors the largest in 2010?	0.21	1	
52	Treemap	The number of unique visitors for Amazon was more than that of Yahoo in 2010.	0.29	0.6	
53		Samsung is nested in the Financial category.	0.38	1	

range of 19 to 79. Everyone had a high school diploma or higher, 40% had a bachelor's degree, and 19% had a master's or a doctoral degree.

Scoring We observed both raw and corrected scores. To calculate the corrected score, we used the correction-for-guessing formula [DE73, Fra88]. Similarly to the original study, the participants were instructed to select the "Skip" option instead of guessing. This setting would influence participants' test-taking strategies and reduce test error caused by guessing, which was a weakness of multiple-choice items [DE73, Fra88]. The correction-for-guessing

formula is defined as follows:

$$CS = R - \frac{W}{C-1}$$

where CS was the corrected score, R was the number of items answered correctly (i.e. raw score), W was the number of items answered incorrectly, and C was the number of choices for an item [LKK16, TTC13]

Results The raw scores of the test takers ranged from 18 to 50 (M = 35, SD = 7.46). The corrected scores ranged from 1 to 48.33 (M = 28.13, SD = 10.05). In addition, we observed the test completion time. The average test completion time was 11 minutes 33 seconds (SD = 2 minutes 17 seconds). The test completion ranged from 5 minutes 45 seconds to 18 minutes 9 seconds. After analyzing the responses from the replication study, we computed the item-total correlation for each item, as shown in Table 1. In the following sections, we validated the items for the Mini-VLAT with the domain experts using the content validity ratio and computed the item-total correlation for each item in the VLAT.

4.1.2. Expert Feedback

In this section, we conducted a content validity evaluation with five domain experts in the visualization community. As the number of experts required for reliable results can be subjective, involving at least five individuals is generally recommended to maintain control over the chance agreement and increase the overall validity of the findings [ZGR*15, PB06]. By having five experts evaluate the relevance of each item to its respective visualization type using a 3-Likert scale, we aimed to ensure that the selected items were appropriate for the Mini-VLAT.

To evaluate content validation, we computed the content validity ratio of each item in the VLAT. The content validity ratio (*CVR*), developed by Lawshe, is a quantitative approach frequently used to assess content validity [Law75, AS14]. *CVR* ranges from -1.0 to 1.0 and indicates expert agreement on how a specific item is required to measure the trait or skill. In our evaluation, we are looking for values above zero, indicating that more than half of the panel members agreed upon an item's relevance to the respective visualization. It helps us in item selection by selecting the items with positive *CVR* and high item-total correlation for each visualization type. The outcome of this evaluation would be the first evidence for validity that test items have the meaning intended when the test was developed [McD13, MH06].

Participants We invited five domain experts in Information Visualization from various locations. They had 5 to 18 years of experience in visualization research and development. All of the experts were from academia.

Procedure We used the following procedure to evaluate with the experts. First, we informed them about the purpose of the study. Then we asked them to rate each item in the VLAT on a 3-Likert scale, *not necessary, useful but not necessary, and essential*, based on their importance to the respective visualization. We calculated CVR for each item using Lawshe's CVR formula based on the number of experts who indicated "essential" in their

responses [AS14].

$$CVR = \frac{n_e - (N/2)}{N/2}$$

n_e is the number of experts indicating an item as "essential," and N is the total number of experts. Table 1 shows the CVR for each item.

4.1.3. Item Selection

Multiple methods exist for selecting short-form items [WLPS11]. For example, Lee et al. [LKK16] suggested creating an effective test by selecting highly discriminating items. This approach considers the *discriminating index* of each item, which is defined as the measure of an item's ability to discriminate between those who scored high on the total test and those who scored low on the total test [HDR02]. An alternative approach uses the *item-total correlation*, defined as the correlation between the scores for a given question to the total scores.

Although there is no consensus for selecting the final items for a test [TTC13], for Mini-VLAT, we chose items with a high item-total correlation. The item-total correlation is more reliable than the items' discriminating index [Bur01]. It makes maximum use of the available information. In contrast, the discriminating index does not use the data for the middle-scoring group of examinees and only incomplete use of the information for the upper and lower groups [Bur01].

After computing the item-total correlation and content validity ratio for each item in the VLAT, we discovered that some of the items (*Item 3*, *Item 8*, *Item 12*, *Item 25*, and *43*) with the highest item-total correlation in their respective visualization type did not have a positive content validity ratio, *CVR*. As a result, we decided to select items with the positive *CVR* and yet high item-total correlation to maintain the validity of the Mini-VLAT, as shown in Table 1.

4.2. Phase II: Piloting and Item Refinement

Once the items were selected, we conducted a pilot study to ensure the test takers easily understood the phrases and vocabulary in the questions. Understanding and assuring the lexical level of potential test takers is critical since the vocabularies and words that make up items might impact test takers' performance [ERAPA99]. During the pilot study, participants reported difficulty answering questions due to the wording and the unfamiliarity of the topic used in the 100% Stacked Bar Chart. We then revised the items based on the feedback received in this study. This ensures that the questions in the final Mini-VLAT are not hard to interpret and not inducing any political bias while answering them.

Pilot Study After selecting the items for the Mini-VLAT, we conducted a pilot study among 15 volunteers via convenience sampling. All of the participants had bachelor's degrees, with 6 of them having completed master's degrees. They consisted of 7 females and 8 males with a self-reported age range of 22 to 59. During the pilot study, 8 participants reported having trouble interpreting some of the questions, and 5 were unfamiliar with the term *exit*

Table 2: *The test items in the Mini-VLAT. The materials used in generating the visualizations (Figure 1) are available at https://osf. io/46rt8/.*

| Item ID	Visualization Type	Question
1	Line Chart	What was the price of a barrel of oil in February 2020?
2	Bar Chart	What is the average internet speed in Japan?
3	Stacked Bar Chart	What is the cost of peanuts in Seoul?
4	100% Stacked Bar Chart	Which country has the lowest proportion of Gold medals?
5	Pie Chart	What is the approximate global smartphone market share of Samsung?
6	Histogram	What distance have customers traveled in the taxi the most?
7	Scatterplot	There is a negative linear relationship between the height and the weight of the 85 males.
8	Area Chart	What was the average price of pount of coffee beans in October 2019?
9	Stacked Area Chart	What was the ratio of girls named "Isla" to girls named "Amelia" in 2012 in the UK?
10	Bubble Chart	Which city's metro system has the largest number of stations?
11	Choropleth Map	In 2020, the unemployment rate for Washington (WA) was higher than that of Wisconsin (WI).
12	Treemap	eBay is nested in the Software category.
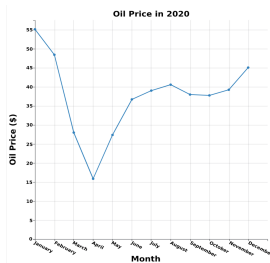
(a) Line Chart
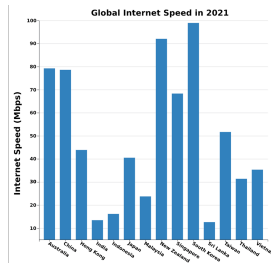
(b) Bar Chart
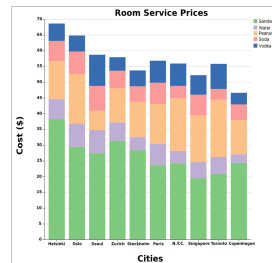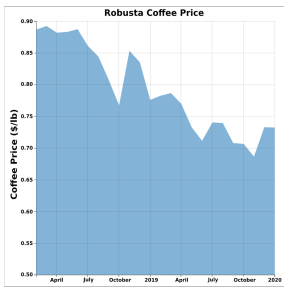
(c) Stacked Bar Chart
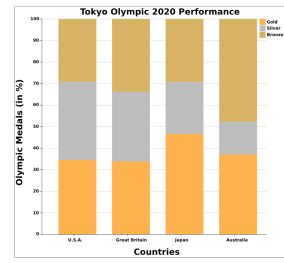
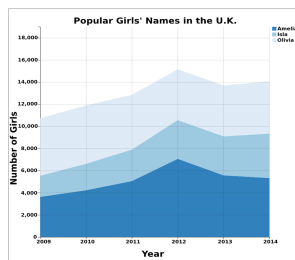(d) 100% Stacked Bar Chart
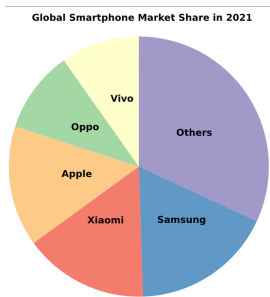
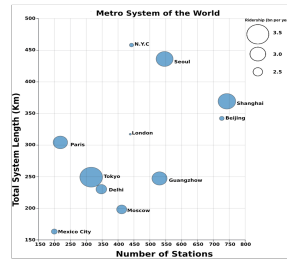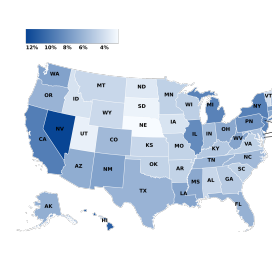(e) Pie Chart
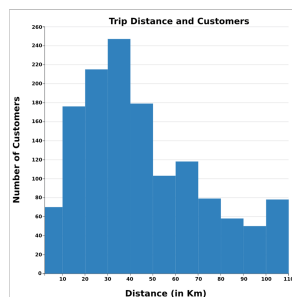
(f) Histogram
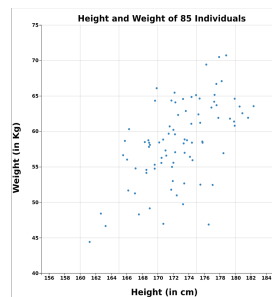
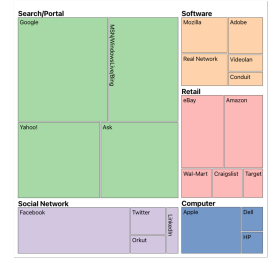(g) Scatterplot

(h) Area Chart

(i) Stacked Area Chart

(j) Bubble Chart

(k) Choropleth Map

(l) Treemap

Figure 1: *The 12 data visualizations used in the Mini-VLAT.*

polls used in the 100% Stacked Bar Chart question comparing approval ratings for the Democrat and Republican political parties in the United States. Some reported that questions were too long or required test takers to read at least twice before answering them.

Revising Questions   To address the feedback received in the Pilot study, we decided to make several changes to the questions in the Mini-VLAT. We updated the dataset used in the VLAT and created the visualizations using D3.js, Figure 1. We also used the colorblind safe colors from *ColorBrewer* in creating the visualizations, making it suitable for color-blind users [CB]. For the question regarding 100% Stacked Bar Chart, we decided to use the Tokyo 2020 Summer Olympics dataset. The visualization used for 100% Stacked Bar Chart in the VLAT can induce bias and influence users' responses due to their prior political beliefs or attitude towards a political party [KLK18]. The updated and finalized version of the questions in the Mini-VLAT is given in Table 2.

4.3. Phase III: Reliability Evaluation

After establishing the validity of the Mini-VLAT, we evaluated the reliability of the Mini-VLAT to measure the internal consistency and stability of the test items. Reliability is the property of observed test scores and the attribute of consistency in a test [Ang53]. A test's reliability can be measured in several ways, including test-retest reliability, parallel test form reliability, and Cronbach's coefficient alpha [Gut45, Lor83, Pet94]. Despite the long-standing and widespread use of reliability measures in test development, there does not appear to be a firm consensus on which measure to use [Sit22]. McDonald's reliability coefficient omega (ω) is another way to calculate the reliability coefficient. It is commonly used in practical situations and is generally preferred over Cronbach's coefficient alpha [HC20, DBB14].

We used the reliability coefficient omega to measure the reliability of the test items in the Mini-VLAT. The result showed that the Mini-VLAT had an acceptable reliability omega of 0.72. This indicates that scores on the test were adequately consistent and were not unduly influenced by random error [Nun67].

4.4. Phase IV: Correlation Between Mini-VLAT & VLAT

An essential component of demonstrating the validity of the short form is showing adequate overlapping variance between a short and full form. The short form is analogous to an alternate form of the full-length measure, and a strong correlation between the two forms is required to support the short form's validity. [SMA00]. In this section, we performed a crowdsourced experiment where the participants took both the short form and the full form. The pre-registration for Phase IV and V and the data obtained from the crowdsourced experiments are available at `https://osf.io/46rt8/`.

Participants   We recruited a total of 30 participants using the same requirements as mentioned in 4.1.1 on Prolific. Since the participants were asked to do the VLAT, we excluded those with color blindness. Out of 30, 17 identified as male, and 13 were female. Their ages ranged from 20 to 65. There were no random clickers in this experiment. Everyone had an education level of high school

and above, 40% of the participant had a bachelor's degree, and 6% of the participants had a master's or a doctoral degree.

Procedure   We randomly assigned participants to either the short form or full-length version of the assessment without revealing the answers to them. Experimenting with 65 items in one sitting can cause participant fatigue and tiredness, resulting in unreliable scores. Therefore, we provided an optional break of 2 minutes before proceeding to the next test.

Result   We recorded the corrected scores using the correction-for-guessing formula for both Mini-VLAT and VLAT. The corrected scores for Mini-VLAT ranged from -1.33 to 10.67 ($M = 5.8$, $SD = 2.9$) and -10.67 to 45.67 ($M = 23.3$, $SD = 13.5$) for VLAT. In addition, we recorded the completion time for both forms. The completion time for the short form ranged from 2 minutes 21 seconds to 4 minutes 54 seconds ($M = 3$ minutes 45 seconds, $SD = 45$ seconds) and 7 minutes 11 seconds to 17 minutes 23 seconds ($M = 12$ minutes 2 seconds, $SD = 2$ minutes 48 seconds) for the full form.

We calculated a Pearson's product-moment correlation coefficient between the Mini-VLAT and VLAT scores. We noticed a strong positive correlation between the two scores ($r = 0.75$, $n = 30$, $p < 0.001$). Since the correlation was significant, it suggests that the Mini-VLAT is at least as valid as the VLAT.

5. PHASE V: TESTING Mini-VLAT'S PREDICTIVE CAPACITY

Finally, we validated whether Mini-VLAT sufficiently measures someone's ability to read and understand visualizations. A widely accepted approach is *concurrent validity*, which involves evaluating the predictive capacity of a test score against an independent criterion or a previously validated measure [GTNK14]. An astute reader may note that this is similar to Phase IV. However, given the overlap in the items between VLAT and mini-VLAT, we opted for the extra step of validating with an independent assessment.

Lee et al. [LKK16] demonstrated a strong positive correlation between visualization literacy and the aptitude for learning an unfamiliar visualization using the VLAT scores and the scores on Parallel Coordinate Plot developed by Kwon and Lee [KL16]. We used a similar procedure to test whether there was any potential correlation between the Mini-VLAT scores and the results of the well-known Parallel Coordinate Test (P-Lite) to determine whether the meaning of the Mini-VLAT scores could be expanded [FDWL22].

Participants   A total of 30 participants were originally recruited through Prolific. We used similar requirements as mentioned in 4.1.1. We ruled out 3 random clickers, and none of the participants were aware of Parallel Coordinate Plot in advance. As a result, a total of 27 participants remained. Out of 27, 15 were females, and 12 were males, with a self-reported age range of 22 to 63. Everyone had an education level of high school and above, 41% of the participant had a bachelor's degree, and 7% of the participants had a master's or a doctoral degree.

Procedure   The experiment was divided into two sections: (1) measuring visualization literacy using the Mini-VLAT and (2) measuring aptitude for learning an unfamiliar visualization. For the

first section, we asked participants to answer the 12 items in the Mini-VLAT. After they finished the first section, we introduced them to six static tutorials on Parallel Coordinate Plot before proceeding to the second section [KL16]. In the second section, we asked the participants to answer the 14 items obtained from the Parallel Coordinate Plot test, P-Lite [FDWL22]. The questions in both sections were randomized.
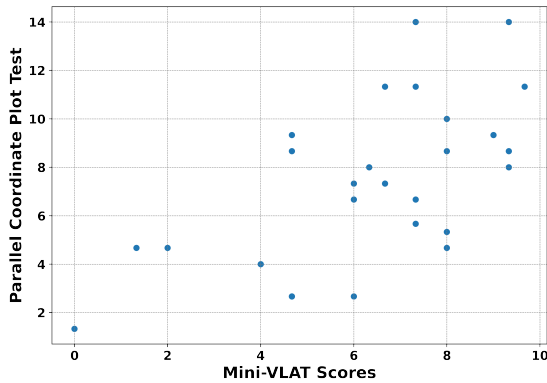

Figure 2: *A scatterplot showing the relationship between the Mini-VLAT's scores and the scores on the Parallel Coordinate Test*

Result   For the Mini-VLAT, we observed the correct scores ranged from 0 to 9.67 (M = 6.36, S = 2.49) and 1.33 to 14 (M = 7.37, S = 3.39) out of 14 for the Parallel Coordinate Test. We calculated Pearson's product-moment correlation coefficient between the Mini-VLAT scores and the post-tutorial scores to evaluate the relationship between visualization literacy and the aptitude for learning. Figure 2 shows a strong positive correlation (r = 0.62, n = 27, p <0.001) between the Mini-VLAT scores and the scores on the aptitude test. This Pearson's correlation coefficient was reasonably high in Psychological Measurement [SBS18]. This outcome demonstrates a strong and favorable correlation between the Mini-VLAT scores and the users' ability to read and understand unfamiliar visualizations. This strong correlation indicates that the Mini-VLAT has utility in measuring someone's ability to read and understand visualizations

6. DISCUSSION

With the increasing use of visualizations, understanding and interpreting the data displayed to formulate accurate observations utilizing visual designs is becoming increasingly important [FL18]. The development of tests, like the one presented in this paper, makes it possible to evaluate users with varying degrees of visualization literacy. However, creating tools/tests to assess visualization literacy is still at the embryonic stage [JBL*21].

6.1. Summary of Results

This study presents Mini-VLAT, a 12-item short form of the Visualization Literacy Assessment Test (VLAT), a reliable and valid instrument for measuring visualization literacy. The Mini-VLAT was found to have good internal consistency reliability (coefficient omega = 0.72) and content validity (average content validity ratio of 0.6), and it strongly correlates with the VLAT. Additionally, the Mini-VLAT was shown to have a strong positive association with users' ability to read and understand unfamiliar visualizations.

The results suggest that Mini-VLAT is a practical and time-effective tool for assessing visualization literacy and can be a suitable substitute for the VLAT. The study also highlights the importance of measuring visualization literacy, as understanding and interpreting visualizations are becoming increasingly important in various fields. While there is no agreement on the definition of visualization literacy, this study focused on the ability to read and interpret visualizations, a key aspect of visualization literacy.

Overall, the study provides a foundation for further research on visualization literacy and encourages the development of multiple instruments or exams to assess different aspects of visualization literacy.

6.2. Toward Measuring Other Dimensions of Visualization Literacy

Several research studies have attempted to define visualization literacy, but there has been no agreement so far [BRBF14, BMBH16, BBG19, LKK16, FJL22]. The ability to interpret is included in every visualization literacy definition and accounts for the majority of the literature [Sol22]. However, some previous works have argued that visualization literacy should be defined more broadly as the ability to reason with graphics: knowing when and how to create a visual representation of data to facilitate information extraction and then knowing how to interpret visual representations to read directly from the data [CRA*18, ARC*17].

It seems unlikely that the research community in the future will agree upon a single definition. As mentioned before, we believe that visualization literacy is a multidimensional construct. This work focused on the user's *ability to read and interpret data visualizations.* We hope that this work will encourage researchers to investigate the topic of visualization literacy more. In the future, we anticipate that multiple instruments or exams can be developed to assess various aspects of visualization literacy.

In contrast to previous works primarily focused on the ability to read and interpret well-formed visualizations, a recent study by Lily et al. [LCK23] aimed to measure people's ability to reason about erroneous or potentially misleading visualizations. The researchers developed a precise definition of misleaders and constructed initial test items using a design space of misleaders and chart types. They then tested the provisional assessment on 497 participants and refined the items using various analysis techniques, including Item Response Theory and qualitative analysis. The final bank of 45 items showed high reliability and provided recommendations for future tests and use cases. This study highlights the importance of measuring visualization literacy from a multidimensional perspective and encourages further research.

7. FUTURE WORK

Data visualizations have been key in disseminating information to the general public. However, there is still much to do in investigating the visual literacy of the citizens to use the visualizations effectively [Len06, BRBF14, LYI*21]. For instance, we need longitudinal studies to understand the ability to comprehend and interpret the visualizations of the general public over a period of time. The Mini-VLAT could assist researchers interested in investigating visualization literacy on a broader scale.

7.1. Visualization Literacy and Inclusivity

During our analysis of previous works on accessing and measuring visualization literacy, we noticed a gap in the demographic information of the participant pools during the user studies. Most papers involved participants living in the United States or Europe [LKK16, PKH21, ARC*17, BMBH16, FDL20].

However, there have been some works involving those who are largely underrepresented in the data visualization literature. Peck et al. [PAEE19] investigated the 42 community members in rural Pennsylvania about their perceptions of data visualization. Sultana et al. [SAR21] explored how the visual communication practices in rural tradition differ from the modern communication practices in rural Bangladesh. The Mini-VLAT can serve as a tool to capture the visualization literacy of those who are notably neglected in visualization-related studies in a shorter duration of time.

In addition, previous work has modified the VLAT to investigate a non-English speaking population [KK21]. Likewise, a potential future work can be to adopt the Mini-VLAT to different languages, allowing the community to uncover possible cultural differences in how people read and interpret visual data. One way to modify Mini-VLAT to different languages is using machine translation services such as *Google Translate*. This can be a quick and cost-effective method to translate the questionnaire into various languages. However, machine translation services may not always provide accurate translations, especially for idiomatic expressions or technical terms. Therefore, involving native speakers of the target language to validate the translation would be crucial to ensure the accuracy and comprehensibility of the Mini-VLAT in different languages. Further Modifying Mini-VLAT to different languages can provide valuable insights into the visualization literacy and interpretation of visual representations among diverse populations.

8. LIMITATION

Although we strictly followed the best practices for creating a short form, several noteworthy limitations exist. For example, the Mini-VLAT is limited to 12 different visualization types. We adopted the visualizations from the VLAT, which surveyed the most frequently occurring visualization types in news outlets and the K-12 curriculum [LKK16]. This assessment does not represent many other common data representations. For example, it would be beneficial to include icon arrays and other statistical charts, often used for medical decision-making [MDF12, OMC12, OKCP19, OPH*15, MOC21]. Other charts, such as radar charts and heat maps, are notably excluded. Thus, future work is necessary to expand the range of vi-

sualization types included in the Mini-VLAT, ensuring that the assessments cover a more comprehensive range of visualizations.

As mentioned by Lee et al. [LKK16], it could be argued that visualization literacy is too complex to assign a number. Laura M. Ahearn [Ahe04] described literacy as multiple rather than unidirectional, focusing on the effect of local conditions on how individuals of the community practice and perceive the significance of literacy. Arneson et al. [AO18] proposed the Visualization Blooming Tool (VBT), an adaption of Bloom's taxonomy, to evaluate scientific visual literacy in undergraduate instruction. Evaluating visual literacy based on this taxonomy might require expert intervention when accessing visual literacy and might not be suitable for large-scale experiments. One of the solutions could be to develop tools or tests that cover the individual stage of the cognitive taxonomy mentioned in the VBT.

One might also argue that the length of the Mini-VLAT is too short, and the exact number of items was discretionary. Nonetheless, the reliability and validity of the shorter 12-item Mini-VLAT are almost as strong as those of the full 53-item VLAT. Since creating tools/tests to access visual literacy is still in the early stage, we believe that having an efficient and validated test is an important step toward investigating visualization literacy [LKK16]. Our findings suggest that the Mini-VLAT is a viable tool for researchers and scholars interested in effectively evaluating visualization literacy. Since this is based on the empirical results presented here, we want to emphasize that researchers should consider the advantages and disadvantages of adopting abbreviated forms [DOBL06]. We believe it is important to use a practical perspective while making this decision, namely that of test participants who are frequently faced with the difficulty of completing long surveys.

9. CONCLUSION

In this paper, we systematically developed an efficient and validated tool, Mini-VLAT, to measure visualization literacy. We also demonstrated the predictive capacity of the Mini-VLAT towards the aptitude for learning on unfamiliar visualizations. We believe that the Mini-VLAT will help the visualization community evaluate visualization literacy on a broader scale and encourage researchers to understand users' visual literacy better and more efficiently. In addition, we have demonstrated how to redesign or shorten a test while maintaining acceptable internal consistency and validity to measure visualization literacy. We intend to build on this work to access the visualization literacy of diverse populations.

10. ACKNOWLEDGMENTS

References

[Ahe04] AHEARN L. M.: Literacy, power, and agency: Love letters and development in nepal. *Language and education 18*, 4 (2004), 305–316. 9

[Ang53] ANGOFF W. H.: Test reliability and effective test length. *Psychometrika 18*, 1 (1953), 1–14. 7

[AO18] ARNESON J. B., OFFERDAHL E. G.: Visual literacy in bloom: Using bloom's taxonomy to support visual learning skills. *CBE—Life Sciences Education 17*, 1 (2018), ar7. 9

[ARC*17] ALPER B., RICHE N. H., CHEVALIER F., BOY J., SEZGIN M.: Visualization literacy at elementary school. In *Proceedings of the 2017 CHI conference on human factors in computing systems* (2017), pp. 5485–5497. 8, 9

[AS14] AYRE C., SCALLY A. J.: Critical values for lawshe's content validity ratio: revisiting the original methods of calculation. *Measurement and evaluation in counseling and development 47*, 1 (2014), 79–86. 5

[Ass74] ASSOCIATION A. P.: Standards for educational & psychological tests. 3

[BBG19] BÖRNER K., BUECKLE A., GINDA M.: Data visualization literacy: Definitions, conceptual frameworks, exercises, and assessments. *Proceedings of the National Academy of Sciences 116*, 6 (2019), 1857–1864. 1, 2, 8

[Ber10] BERTIN J.: *Semiology of graphics: Diagrams, networks, maps.* Esri Press, 2010. 2

[BHH*21] BACH B., HURON S., HINRICHS U., ROBERTS J. C., CARPENDALE S.: Special issue on visualization teaching and literacy. *IEEE Computer Graphics and Applications 41*, 06 (2021), 13–14. 1

[BMBH16] BÖRNER K., MALTESE A., BALLIET R. N., HEIMLICH J.: Investigating aspects of data visualization literacy using 20 information visualizations and 273 science museum visitors. *Information Visualization 15*, 3 (2016), 198–213. 2, 8, 9

[BRBF14] BOY J., RENSINK R. A., BERTINI E., FEKETE J.-D.: A principled way of assessing visualization literacy. *IEEE transactions on visualization and computer graphics 20*, 12 (2014), 1963–1972. 1, 2, 8, 9

[Bur01] BURTON R. F.: Do item-discrimination indices really help us to improve our tests? *Assessment & Evaluation in Higher Education 26*, 3 (2001), 213–220. 5

[Car92] CARSWELL C. M.: Choosing specifiers: An evaluation of the basic tasks model of graphical perception. *Human factors 34*, 5 (1992), 535–554. 1, 2

[CB] CYNTHIA BREWER M. H.: Colorbrewer 2.0. Accessed: 2022-11-22. URL: https://colorbrewer2.org/. 7

[CRA*18] CHEVALIER F., RICHE N. H., ALPER B., PLAISANT C., BOY J., ELMQVIST N.: Observations and reflections on visualization literacy in elementary school. *IEEE computer graphics and applications 38*, 3 (2018), 21–29. 8

[CSP96] COHEN R. J., SWERDLIK M. E., PHILLIPS S. M.: *Psychological testing and assessment: An introduction to tests and measurement.* Mayfield Publishing Co, 1996. 1

[CW98] CANIVEZ G. L., WATKINS M. W.: Long-term stability of the wechsler intelligence scale for children—third edition. *Psychological Assessment 10*, 3 (1998), 285. 3

[DBB14] DUNN T. J., BAGULEY T., BRUNSDEN V.: From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British journal of psychology 105*, 3 (2014), 399–412. 7

[DE73] DIAMOND J., EVANS W.: The correction for guessing. *Review of educational research 43*, 2 (1973), 181–191. 4

[DOBL06] DONNELLAN M. B., OSWALD F. L., BAIRD B. M., LUCAS R. E.: The mini-ipip scales: tiny-yet-effective measures of the big five factors of personality. *Psychological assessment 18*, 2 (2006), 192. 3, 9

[ERAPA99] EDUCATIONAL RESEARCH ASSOCIATION A., PSYCHOLOGICAL ASSOCIATION A.: National council of measurement in education.(1999). *Standards for educational and psychological testing* (1999), 63–74. 5

[FDL20] FIRAT E. E., DENISOVA A., LARAMEE R. S.: Treemap literacy: A classroom-based investigation. In *Eurographics Proceedings* (2020). 2, 9

[FDWL22] FIRAT E. E., DENISOVA A., WILSON M. L., LARAMEE R. S.: P-lite: A study of parallel coordinate plot literacy. *Visual Informatics* (2022). 2, 3, 7, 8

[FJL22] FIRAT E. E., JOSHI A., LARAMEE R. S.: Interactive visualization literacy: The state-of-the-art. *Information Visualization 21*, 3 (2022), 285–310. 8

[FL18] FIRAT E. E., LARAMEE R. S.: Towards a survey of interactive visualization for education. In *Proc. Conf. Comput. Graph. Vis. Comput.* (2018), pp. 91–101. 8

[Fra88] FRARY R. B.: Formula scoring of multiple-choice tests (correction for guessing). *Educational measurement: Issues and practice 7*, 2 (1988), 33–38. 4

[GGR11] GALESIC M., GARCIA-RETAMERO R.: Graph literacy: A cross-cultural comparison. *Medical decision making 31*, 3 (2011), 444–457. 2

[GTNK14] GRANPEESHEH D., TARBOX J., NAJDOWSKI A., KORNACK J.: *Evidence-based treatment for children with autism: the CARD model.* Elsevier, 2014. 7

[Gut45] GUTTMAN L.: A basis for analyzing test-retest reliability. *Psychometrika 10*, 4 (1945), 255–282. 7

[HC20] HAYES A. F., COUTTS J. J.: Use omega rather than cronbach's alpha for estimating reliability. but.... *Communication Methods and Measures 14*, 1 (2020), 1–24. 7

[HDR02] HALADYNA T. M., DOWNING S. M., RODRIGUEZ M. C.: A review of multiple-choice item-writing guidelines for classroom assessment. *Applied measurement in education 15*, 3 (2002), 309–333. 5

[JBL*21] JOSHI A., BÖRNER K., LARAMEE R. S., HARRISON L., FIRAT E. E., KWON B. C.: Visualization literacy for general audiences-can we make a difference?, 2021. http://www.cs.nott.ac.uk/blaramee/research/panel/joshi21visualization.pdf. 1, 8

[KBC01] KOEDINGER K. R., BAKER R., CORBETT A.: Toward a model of learning data representations. In *Proceedings of the Twenty-Third Annual Conference of the Cognitive Science Society, Mahwah, NJ: Erlbaum* (2001), pp. 45–50. 2

[KK21] KIM H.-N., KIM S.-H.: Development on korean visualization literacy assessment test (K-VLAT) and research trend analysis. *Journal of the Korea Institute of Information and Communication Engineering 25*, 11 (2021), 1696–1707. 9

[KL16] KWON B. C., LEE B.: A comparative evaluation on online learning approaches using parallel coordinate visualization. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (2016), pp. 993–997. 7, 8

[KLK18] KONG H.-K., LIU Z., KARAHALIOS K.: Frames and slants in titles of visualizations on controversial topics. In *Proceedings of the 2018 CHI conference on human factors in computing systems* (2018), pp. 1–12. 7

[Law75] LAWSHE C. H.: A quantitative approach to content validity. *Personnel psychology 28*, 4 (1975), 563–575. 5

[LCK23] LILY W. G., CUI Y., KAY M.: Calvi: Critical thinking assessment for literacy in visualizations. 8

[LCO20] LIU Z., CROUSER R. J., OTTLEY A.: Survey on individual differences in visualization. In *Computer Graphics Forum* (2020), vol. 39, Wiley Online Library, pp. 693–712. 1

[LD96] LAFRENIERE P. J., DUMAS J. E.: Social competence and behavior evaluation in children ages 3 to 6 years: The short form (scbe-30). *Psychological assessment 8*, 4 (1996), 369. 3

[Len06] LENGLER R.: Identifying the competencies of'visual literacy'-a prerequisite for knowledge visualization. In *Tenth International Conference on Information Visualisation (IV'06)* (2006), IEEE, pp. 232–236. 9

[LKK16] LEE S., KIM S.-H., KWON B. C.: VLAT: Development of a visualization literacy assessment test. *IEEE transactions on visualization and computer graphics 23*, 1 (2016), 551–560. 1, 2, 3, 5, 7, 8, 9

[LKY*19] LEE S., KWON B. C., YANG J., LEE B. C., KIM S.-H.: The correlation between users' cognitive characteristics and visualization literacy. *Applied Sciences 9*, 3 (2019), 488. 2, 3

[Lor83] LORD F. M.: Unbiased estimators of ability parameters, of their variance, and of their parallel-forms reliability. *Psychometrika 48*, 2 (1983), 233–245. 7

[LYI*21] LEE C., YANG T., INCHOCO G. D., JONES G. M., SATYANARAYAN A.: Viral visualizations: How coronavirus skeptics use orthodox data practices to promote unorthodox science online. In *Proceedings of the 2021 CHI conference on human factors in computing systems* (2021), pp. 1–18. 9

[McD13] MCDONALD R. P.: *Test theory: A unified treatment*. Taylor and Francis, 2013. 5

[MDF12] MICALLEF L., DRAGICEVIC P., FEKETE J.-D.: Assessing the effect of visualizations on bayesian reasoning through crowdsourcing. *IEEE transactions on visualization and computer graphics 18*, 12 (2012), 2536–2545. 9

[Mel87] MELZACK R.: The short-form mcgill pain questionnaire. *Pain 30*, 2 (1987), 191–197. 3

[MH06] MISLEVY R. J., HAERTEL G. D.: Implications of evidence-centered design for educational testing. *Educational measurement: issues and practice 25*, 4 (2006), 6–20. 3, 5

[MOC21] MOSCA A., OTTLEY A., CHANG R.: Does interaction improve bayesian reasoning with visualization? In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (2021), pp. 1–14. 9

[MVH17] MANSOOR H., VASU K., HARRISON L.: Linking performance on graphical perception tasks to visualization literacy. URL: https://web.cs.wpi.edu/~hmansoor/vizliteracy_poster.pdf. 2, 3

[Nun67] NUNNALLY J. C.: *Psychometric theory*. McGraw-Hill, 1967. 7

[OKCP19] OTTLEY A., KASZOWSKA A., CROUSER R. J., PECK E. M.: The curious case of combining text and visualization. *EuroVis 2019-Short Papers* (2019). 9

[OMC12] OTTLEY A., METEVIER B., CHANG R.: Visually communicating bayesian statistics to laypersons. 2012. 9

[OPH*15] OTTLEY A., PECK E. M., HARRISON L. T., AFERGAN D., ZIEMKIEWICZ C., TAYLOR H. A., HAN P. K., CHANG R.: Improving bayesian reasoning: The effects of phrasing, visualization, and spatial ability. *IEEE transactions on visualization and computer graphics 22*, 1 (2015), 529–538. 9

[Ott20] OTTLEY A.: Adaptive and personalized visualization. *Synthesis Lectures on Visualization 7*, 1 (2020), 1–117. 1

[PAEE19] PECK E. M., AYUSO S. E., EL-ETR O.: Data is personal: Attitudes and perceptions of data visualization in rural pennsylvania. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (2019), pp. 1–12. 1, 9

[PB06] POLIT D. F., BECK C. T.: The content validity index: are you sure you know what's being reported? critique and recommendations. *Research in nursing & health 29*, 5 (2006), 489–497. 5

[Pet94] PETERSON R. A.: A meta-analysis of cronbach's coefficient alpha. *Journal of consumer research 21*, 2 (1994), 381–391. 7

[PKH21] PEPPLER K., KEUNE A., HAN A.: Cultivating data visualization literacy in museums. *Information and Learning Sciences 122*, 1/2 (2021), 1–16. 1, 9

[PYH*12] PECK E. M., YUKSEL B. F., HARRISON L., OTTLEY A., CHANG R.: Towards a 3-dimensional model of individual cognitive differences: Position paper. In *Proceedings of the 2012 BELIV Workshop: Beyond Time and Errors-Novel Evaluation Methods for Visualization* (2012), pp. 1–6. 1

[SAR21] SULTANA S., AHMED S. I., RZESZOTARSKI J. M.: Seeing in context: Traditional visual communication practices in rural bangladesh. *Proceedings of the ACM on Human-Computer Interaction 4*, CSCW3 (2021), 1–31. 9

[SBS18] SCHOBER P., BOER C., SCHWARTE L. A.: Correlation coefficients: appropriate use and interpretation. *Anesthesia & Analgesia 126*, 5 (2018), 1763–1768. 8

[Sit22] SITARENIOS G.: Short versions of tests: Best practices and potential pitfalls. *Journal of Pediatric Neuropsychology* (2022), 1–15. 1, 3, 7

[SMA00] SMITH G. T., MCCARTHY D. M., ANDERSON K. G.: On the sins of short-form development. *Psychological assessment 12*, 1 (2000), 102. 1, 3, 7

[Sol22] SOLEN M.: Scoping the future of visualization literacy: A review. URL: osf.io/eypgm. 8

[SSBS02] STANTON J. M., SINAR E. F., BALZER W. K., SMITH P. C.: Issues and strategies for reducing the length of self-report scales. *Personnel Psychology 55*, 1 (2002), 167–194. 3

[TTC13] THORNDIKE R. M., THORNDIKE-CHRIST T. M.: *Measurement and evaluation in psychology and education*. Pearson Higher Ed, 2013. 5

[Wai92] WAINER H.: Understanding graphs and tables. *Educational researcher 21*, 1 (1992), 14–23. 2

[WLPS11] WIDAMAN K. F., LITTLE T. D., PREACHER K. J., SAWALANI G. M.: *On creating and using short forms of scales in secondary research*. American Psychological Association, 2011, pp. 39–61. 5

[Yi12] YI J. S.: Implications of individual differences on evaluating information visualization techniques. *International Journal of Human-Computer Studies 45*, 6 (2012), 619–637. 1

[ZGR*15] ZAMANZADEH V., GHAHRAMANIAN A., RASSOULI M., ABBASZADEH A., ALAVI-MAJD H., NIKANFAR A. R.: Design and implementation content validity study: development of an instrument for measuring patient-centered communication. *Journal of caring sciences 4*, 2 (2015), 165–178. 5