What University Students Learn In Visualization Classes

Maryam Hedayati and Matthew Kay

Abstract— As a step towards improving visualization literacy, this work investigates how students approach reading visualizations differently after taking a university-level visualization course. We asked students to verbally walk through their process of making sense of unfamiliar visualizations, and conducted a qualitative analysis of these walkthroughs. Our qualitative analysis found that after taking a visualization course, students engaged with visualizations in more sophisticated ways: they were more likely to exhibit *design empathy* by thinking critically about the tradeoffs behind why a chart was designed in a particular way, and were better able to *deconstruct* a chart to make sense of it. We also gave students a quantitative assessment of visualization literacy and found no evidence of scores improving after the class, likely because the test we used focused on a different set of skills than those emphasized in visualization classes. While current measurement instruments for visualization literacy are useful, we propose developing standardized assessments for additional aspects of visualization literacy, such as deconstruction and design empathy. We also suggest that these additional aspects could be incorporated more explicitly in visualization courses. All supplemental materials are available at `https://osf.io/w5pum/`.

Index Terms—visualization literacy, visualization pedagogy, graph comprehension, visualization expertise

1 Introduction

We regularly encounter and use data visualizations across all aspects of our lives: from consulting a weather forecast, to checking financial statements, to making health decisions. Visualizations can help us synthesize the vast amount of information that we are exposed to—so long as we are able to read and interpret them correctly. It is therefore increasingly important for people to learn how to effectively understand and use data visualizations. People's ability to work with visualizations has been dubbed their *visualization literacy*, and has been defined as "the ability and skill to read and interpret visually represented data in and to extract information from data visualizations" [35].

As visualization researchers and educators, it is crucial for us to understand how to measure and improve people's ability to read visualizations, including unfamiliar ones they may encounter in the world. But to improve it, it is first necessary to understand how expertise in visualization reading develops. University-level visualization classes offer a setting where students explicitly focus on data visualization for an extended period of time. We took advantage of this existing setting to study how students' visualization reading abilities develop over time. While our study was designed to understand how visualization expertise develops, our results also have possible implications for improving visualization courses and visualization literacy assessments.

We specifically studied how university students read visualizations differently before and after taking a visualization course. We interviewed students enrolled in university-level visualization courses over two sessions: one session at the beginning of their course, and one at the end. In each session, we gained insight into participants' graph comprehension processes by having them verbally walk through their process of making sense of two unfamiliar visualization types. Our goal was to understand how participants make sense of visualizations they have not previously seen, rather than their ability to read a familiar visualization.

We qualitatively analyzed participants' verbal walkthroughs of unfamiliar visualizations, and found that both before and after the class, participants' sensemaking processes followed many of the same steps, some of which were reflected in existing models of graph comprehension. However, there were several differences in the way students made sense of visualizations after taking a visualization course. After a visualization course, students were better able to **deconstruct** the visualizations they were interpreting. For example, participants were more likely to focus on specific details when making sense of and explaining a visualization, and used prior knowledge about visualization types to deconstruct the visualizations they were interpreting. They also were more likely to exhibit **design empathy**—understanding and empathizing with the designer's perspective when reading visualizations. They discussed choices made in the design of the visualization, focusing less on what each visual feature represented and more on why the feature was chosen, demonstrating awareness of design tradeoffs.

Despite qualitative evidence that participants read and understood visualizations in more sophisticated ways after taking a visualization course, we found no evidence of quantitative improvements on a visualization literacy assessment [35]. The qualitative improvements that we found reflected important aspects of literacy not captured by existing quantitative visualization literacy scales. While current measurement instruments are valuable, visualization educators and researchers would benefit from validated assessments for these other skills, including the ability to deconstruct visualizations, critique visualizations, or understand design tradeoffs. In addition, while implicitly covered within other topics, visualization educators could explicitly emphasize learning goals related to two of our findings: the ability to deconstruct visualizations and the ability to empathize with the designer when reading visualizations. More systematic measurement of these skills could also be used to evaluate and refine visualization teaching methods, feeding back into improved learning goals and educational material in visualization classes.

2 Related Work

2.1 Visualization Literacy

Although traditional definitions of literacy refer to the ability to read and write text, other literacies have been introduced to reflect the different ways that people communicate [13]. In addition to textual literacy and numerical literacy, many additional forms of literacy have been proposed, including media literacy and information literacy [53]. In this category is **visualization literacy**, which has been used as a way of measuring the ability to read data visualizations. Definitions of visualization literacy include features such as the ability to "translate questions specified in the data domain into visual queries in the visual domain", to "interpret visual patterns in the visual domain as properties in the data domain" [7], to "read and interpret visually represented data" and to "extract information from data visualizations [35]. Although some definitions of visualization literacy include the ability to create new visualizations [10], most focus primarily on the ability to understand and work with existing visualizations [26].

Several measures of visualization literacy have been proposed for use cases such as assessing ability levels in educational contexts and assessing the target audience for a particular visualization [7, 35]. These tests include Boy et al.'s [7] assessments of people's ability to

• *Maryam Hedayati and Matthew Kay are with Northwestern University. E-mail: {maryam.hedayati | mjskay}@northwestern.edu .*

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

read line graphs, bar charts, and scatter plots, and the Visualization Literacy Assessment Test (VLAT) [35], which includes multiple-choice questions about 12 different visualizations, based on data visualization tasks such as "Retrieve Value", "Determine Range", and "Find Correlations/Trends". Other visualization literacy measures typically also use multiple-choice and short answer questions based on lower level visualization tasks [22, 56]. While these tests are useful for measuring low-level tasks, they do not provide a full picture of someone's understanding of a visualization [9]. We discuss this further in Sec. 6.3.

Some assessments add other components to visualization literacy. Ge et al.'s CALVI [24] measures critical thinking by incorporating questions that require reasoning about incorrect visualizations. Börner et al. [10] emphasize the importance of assessing the construction of data visualizations, and propose using a rubric to do so, which is typically how construction is assessed in visualization classes [5, 43]. Berg & Smith [4] assessed the construction and interpretation of line charts by asking participants to fill in a pre-labeled graph based on a provided scenario, but find that multiple-choice questions may not be a valid measure of graphing abilities given the disparities between participant answers to multiple-choice questions and interview questions. Some assessments ask questions about familiarity or perceived competence with particular visualization types, a construct sometimes referred to as subjective visualization literacy [11, 23, 31, 39, 44].

Beyond explicit assessments, there are other approaches to studying visualization literacy. Börner et al. [11] interviewed museum visitors about common visualization types. They asked participants to name the visualizations, state their familiarity with them, and talk about how they would decode them. Because of the constraints of the museum setting, they did not go in depth on how people would read the visualizations, and focused primarily on common chart types [11]. Other work looks at how visualization literacy is taught in early grades by analyzing textbooks, surveying teachers, and introducing tools to help improve visualization literacy [1, 14]. Firat et al. provide a useful survey of prior research studying visualization literacy, with a focus on interactive visualizations [19].

2.2 Graph Comprehension

While *visualization literacy* is typically operationalized as a quantification of a person's **ability** to correctly read graphs and charts, *graph comprehension* refers to the cognitive **processes** that take place when somebody encounters a graph or chart [47]. Graph comprehension specifically focuses on the processes that take place as someone looks at a graph rather than only whether they are accurately interpreting it. Some models describe graph comprehension as an interaction between visual features, perceptual processes, and graph schema [47]. Freedman & Shah [20] propose the construction-integration model, in which visual features and domain knowledge are combined with interpretation propositions to create an understanding of the graph. Other models of graph comprehension include Carpenter & Shah's model [12], which describes graph comprehension as an iterative process, and Trickett & Trafton's model [54], which incorporates spatial cognition.

Some models of graph comprehension specifically focus on how people understand unfamiliar visualizations, such as the NOVIS model by Lee et al. which breaks the process of understanding a visualization into five steps [34]. VLAT scores are correlated with the ability to learn how to use an unfamiliar visualization, and therefore the ability to successfully understand a new visualization type seems to be correlated with visualization literacy levels [35]. However, we do not know how this ability to interpret unfamiliar visualizations develops, motivating our study of students' interpretations of unfamiliar visualizations before and after taking a visualization class.

3 Study Design

Our goal in designing this study was to understand how students approached visualizations differently after taking a visualization course. We specifically focused on university-level visualization courses, as visualization tends to be integrated into other subjects during K-12 education, and we wanted to focus on a setting in which students explicitly focus on visualization for an extended period of time. We asked students to talk aloud as they walked through unfamiliar visualization types to see how

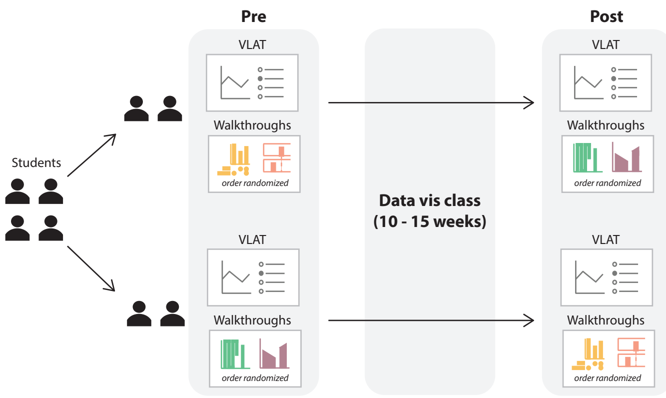

Fig. 1: Overview of the study design. Participants were randomly assigned to one of two groups. During each study session, they completed the VLAT and a walkthrough of two unfamiliar visualizations. The visualizations they saw in each session were determined by the group they were assigned to.

they made sense of them. By doing this before and after the course, we were able to look at how the same set of participants engaged with visualizations differently after gaining experience working with visualizations.

This study was conducted in two sessions: PRE and POST. Each participant completed the first session (PRE) during the first week of the 10-week or 15-week data visualization course they were enrolled in, and the second session (POST) during the final week of the course. The period of time between the two sessions helped reduce the extent to which participants' responses in the second session were influenced by the questions asked in the initial session. In each session, participants did a **verbal walkthrough of two unfamiliar visualizations**, which were chosen to be somewhat similar to visualizations that students may have previously seen, but different enough that they would need to figure out how to read them. We used walkthroughs to gain insight into graph comprehension because similar tasks have been used to investigate how novices make sense of unfamiliar charts [34]. Participants also completed a **shortened version of the Visualization Literacy Assessment Test** (see Sec. 3.2) [35]. The study components are illustrated in Fig. 1.

This work was part of a larger interview study conducted in Fall 2021, which was approved by Northwestern University's Institutional Review Board. Each full session was conducted via Zoom and lasted around 60 minutes. Participants were paid \$15 for each session, with a \$20 bonus for completing both sessions. The interview questions from the remainder of the interview are included in supplemental materials. Our focus in this paper is participants' graph comprehension processes, and we therefore analyze participants' initial walkthroughs of the two unfamiliar visualizations, which took around 5 minutes per visualization. The VLAT portion of the study took around 15 minutes.

3.1 Walkthroughs of Unfamiliar Visualizations

In the walkthrough portion of the study, which was inspired by the task used by Lee et al. [34], the interviewer (the first author) asked participants to "talk aloud about what [they were] thinking as [they were] looking at [the chart] and making sense of it", telling them that our goal was "to see how easy the charts [were] to understand and [their] strategies for thinking about new visualization types". Participants were reminded to continue to think aloud if they were silent for a long period of time and were prompted for clarification when it was not clear which part of the visualization they were referring to. We provided clarifications for abbreviations they were unsure about, but did not provide other explanations of the chart. The full instructions are provided in supplemental materials.

3.1.1 Unfamiliar Visualizations

We selected unfamiliar visualizations that we anticipated most participants had not previously been exposed to. We therefore excluded visualizations that may be less common in the wild but which are commonly taught in visualization courses, such as treemaps, bump charts, and parallel coordinates plots, all of which appear in visualization textbooks [42]. We also did not want the visualizations to be drastically different from the types of visualizations people are typically exposed to, so we selected charts with features of common chart types, but fundamental differences in how the chart is structured. For example, WRAPPED BAR CHART works similarly to a bar chart, but the bars wrap around once they reach a certain threshold.

We found potential unfamiliar visualizations by brainstorming visualization types that we thought students might not have encountered, asking colleagues for suggestions, and searching Google Scholar, Xenographics [33], and other websites. We ultimately chose four unfamiliar visualization types that best fit our criteria: WRAPPED BAR CHART [29], UPSET [37], MARIMEKKO SLOPE CHART [40], and SCALE-STACK BAR CHART [27]. We verified that charts were unfamiliar by asking participants if they had previously seen each chart type. If they answered yes, we asked them to elaborate further on what they had previously seen, and determined whether they were describing the given chart type. For example, P3-POST reported having previously seen SCALE-STACK BAR CHART, but we realized based on follow-up questions that they had not. When asked where they had seen it, they said "I saw it in the link you sent me.", referring to the VLAT link, which contains a stacked bar chart, but not SCALE-STACK BAR CHART. Only one of the participants (whose data was then excluded) had previously seen any of the chart types.

The particular visualizations we used were drawn from academic papers and blog posts. We added titles to all charts, removed descriptions, and made modifications based on pilot studies to make the charts easier to understand without descriptions. The original versions of the visualizations and the modifications made are listed in the supplemental materials. Here, we briefly describe each of our chosen visualizations and their sources.

WRAPPED BAR CHART (Fig. 2a) aims to visualize data with a large range of values. It is set up like a bar chart, but with bars that wrap around when they reach a threshold [29]. The version used here, from a blog post [28], shows the number of COVID-19 cases in several countries.

UPSET (Fig. 2b) serves as an alternative to a Venn Diagram, and visualizes set intersections. It contains two bar charts, with a combination matrix identifying the intersections depicted in the bar chart above it [37]. The version used here is an implementation [17] using the UpSetR R package [15], which displays results from a survey in Toronto identifying transportation needs of senior citizens.

MARIMEKKO SLOPE CHART (Fig. 2c) uses a bar chart to visualize categorical data, but with the width of each bar scaled by the size of each category and the slope on each bar showing the change in some variable for each group over two time periods. The version used here is from a blog post and shows trends in median income for different educational attainment groups, as well as the population sizes of each group [40].

SCALE-STACK BAR CHART (Fig. 2d) displays stacked bar charts of the same data on multiple scales to visualize data with a large range of values. The version we used, from Hlawatsch et al. [27], is a visualization of the population of several countries with vastly different populations, and the breakdown of age demographics for each country.

We grouped the visualizations into pairs: WRAPPED BAR CHART with MARIMEKKO SLOPE CHART, and UPSET with SCALE-STACK BAR CHART to make sure the topics of the visualizations in each pair were sufficiently different from each other. The session in which each participant saw each pair was counterbalanced across participants, and the order of the two visualizations within each session was randomized, i.e. each pair of visualizations was seen by half of participants in PRE, and by the remaining half of participants in POST.

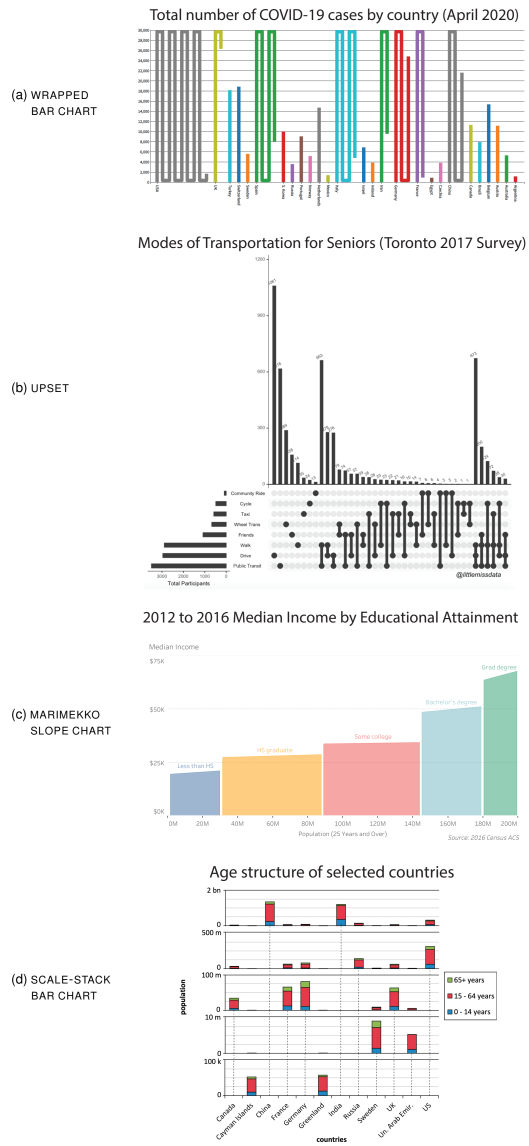

Fig. 2: Unfamiliar visualizations used for the walkthrough portions of the study. Higher resolution versions are available in the supplement.

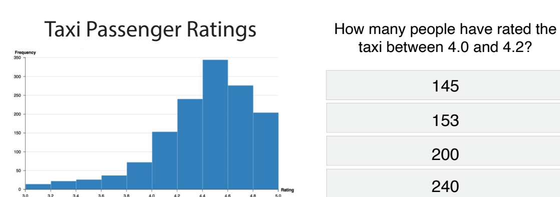

Fig. 3: Sample question from VLAT.

3.2 Modified VLAT

The VLAT consists of 53 multiple-choice questions across 12 visualization types (see sample question in Fig. 3), and the intended audience is non-expert users in data visualizations [35] .

To reduce the amount of time spent on this test,[1] and because we expected university students (particularly those who chose to enroll in a visualization course) to be more skilled at answering visualization questions than the target audience of VLAT, we had participants answer a subset of the VLAT questions. Lee et al. report an item difficulty index for each item in the test [35], and we used these values to select the three most difficult questions for each visualization type, because items with a high difficulty index will provide more information about a more skilled person [3]. Our modified version of the VLAT therefore consisted of 36 total questions.[2] Our participants performed well on the modified

[1]The original VLAT takes around 23 minutes (53 questions), and our modified version takes around 15 minutes (36 questions)

[2]After we conducted our study, two papers were published with shortened

VLAT containing only the most difficult questions, which indicates that our strategy for shortening the assessment was appropriate—the less difficult questions would likely not have provided additional information about our participants.

3.3 Recruitment

3.3.1 Visualization Courses

We recruited participants from university courses to see how students change as they develop visualization expertise. Our goal was to recruit from upper-year visualization classes that were similar enough to each other that we could draw conclusions about the effects of being in such a class, while recruiting from multiple classes so that our findings would not be tied to the idiosyncrasies of one class or instructor.

We chose the courses to recruit from by listing universities with faculty members in the visualization community and looking at their course catalogs for visualization courses in the term we conducted the study (Fall 2021). We also searched online for visualization courses. We reached out to instructors, and shared our recruitment material with those who were willing to participate.

All the courses were at R1 universities in the USA, were at a similar level, and were targeted towards upper-level undergraduates. In Tab. 1, we provide further details about the visualization courses. Based on the course information we have access to (syllabi for C1, C2, C3, C5 and a detailed course schedule for C4), we see many similarities across the classes. In fact, several used course materials derived from other courses we recruited from; specifically, C1, C3, and C5 had notable overlap in course content. All five courses included the following topics: what makes a visualization effective, data types, visual encodings, storytelling, visualization design, and critique. All courses also introduced tools for implementing visualizations, including Tableau (C1, C4, C5), D3 (C1, C3, C5), Vega-lite (C3), ggplot (C2), Shiny (C2), and Python (C4). Some courses addressed specific visualization types, including maps (C1, C2, C3, C5), network visualizations (C1, C3, C5), and text visualizations (C1, C5). Other course topics included perception (C1, C3, C5), color (C1, C3, C4), exploratory data analysis (C1, C2, C3, C5), uncertainty (C1, C3, C5), interaction (C1, C3, C4, C5), grammar of graphics (C2, C4), and deceptive visualizations (C3, C4, C5).

3.3.2 Participants

We recruited students from the courses described in Sec. 3.3.1 and paid students for their participation in each session, with a bonus for completing both sessions. The PRE session was completed by 23 participants in the first week of class, and 19 of those participants also completed the POST session in the final week of class. Of those 19, one participant was excluded for having dropped the class, one was excluded for having previously seen one of the unfamiliar visualizations, and one was excluded for multi-tasking and not paying attention during the POST session. These exclusion decisions were made prior to looking at the VLAT scores and POST session transcripts and resulted in data from 16 students. In addition, 4 pilot participants completed the study in previous terms. Data from pilot participants and excluded participants were used for developing qualitative codes, but not for the final analyses. Of the 16 participants included in the final analysis, 2 participants were graduate students, 10 were fourth-year undergraduates, 2 were third-year undergraduates, and 2 were second-year undergraduates. The participants' ages ranged from 19 to 26 years old. The majority of the participants (13 out of 16) had previously taken a course in statistics (6 at the high school level, 6 at the university level, and 1 unspecified), 5 had taken a course in data science, and 2 had taken a class in data visualization. Outside of their classes, 4 participants reported having used data visualization in a research context, and 2 participants had completed internships incorporating data visualization. We did not exclude participants with prior visualization experience, because it is common for students with prior experience to enroll in visualization classes and we wanted an authentic sample of visualization students. Because of the longitudinal nature of the study, participants were being compared to themselves in the analysis, which allowed us to focus on how participants changed from their individual baselines.

versions of VLAT: Pandey & Ottley's mini-VLAT [45], and Cui et al.'s A-VLAT [16]. It is difficult to directly compare our version to theirs given that mini-VLAT made changes to the scenarios and graphs, and A-VLAT is not a static assessment. However, these works corroborate the need for a shorter version of VLAT.

4 Data Analysis Methods

We analyzed the visualization walkthroughs using a qualitative approach. We analyzed each walkthrough separately to allow for more granularity—if a participant used the same code to talk about multiple visualizations, we wanted this reflected in our data. In addition, because the visualizations we used were all somewhat idiosyncratic, we wanted our unit of analysis to be the combination of participant and visualization (our counterbalancing allows us to compare these code counts between PRE and POST). We also looked at descriptive statistics of VLAT scores.

4.1 Walkthrough qualitative coding

Initial code development. We used grounded theory [41] to begin our coding process by breaking our pilot walkthroughs into sentences and phrases and clustering these based on themes. Because we wanted to compare how people talked about multiple visualization types, we looked for themes that were not specific to one visualization type. We primarily used process coding, which uses gerunds to capture actions in the data [41], to come up with our initial codes. Our initial code development included a set of higher-level codes, some of which had sub-codes. We use the notation *code.subcode* to depict these relationships—the code before the dot is the primary code, and the code after the dot is a sub-code of the primary code. The initial code development was primarily done by the first author, with discussions and support from the last author.

Refinement with data. After developing an initial set of codes based on the pilot data, we added in transcripts from a few more interviews to see if they fit well into our coding scheme. We asked members of our lab (a visualization research lab) to look at parts of the data, both with and without the coding scheme, to see if they noticed themes that we did not or if they had feedback on the coding scheme. The first author also went through some individual transcripts line by line, rephrasing each section based on what it seemed like participants were doing. We used the additional data, the feedback from colleagues, and the rephrased transcripts to refine the coding scheme.

Refinement with graph comprehension models. After creating an initial coding scheme and refining it using an inductive approach [18], we also refined the codes using a deductive approach based on existing models of graph comprehension. We looked at models including Roth & Bowen's semiotic model [49], Freedman & Shah's construction-integration model [20], and Lee et al.'s NOVIS model [34], and modified our coding scheme to incorporate aspects of those models. For example, participants stating their initial interpretation of a chart was first categorized as *purpose.goal*,[3] because a proposed interpretation could be thought of as a statement about a goal of the chart, even if the participant is not confident in their answer. However, based on the NOVIS model [34], which includes the idea of building a "frame" (a possible interpretation of the chart), we recategorized proposed interpretations as a new sub-code of *explain*: *explain.propose*. Two other sub-codes were also added to *explain* to reflect the NOVIS model and the idea of building and testing frames: *explain.verify* (verifying the correctness of their current interpretation), and *explain.modify* (modifying their current interpretation). *Modify* was previously a separate code, but was re-categorized as a sub-code of *explain* so that all the codes about creating and modifying interpretations would be in one category. We also clarified the distinction between prior knowledge about graphs (*priorKnowledge.otherCharts*) and other prior knowledge (*priorKnowledge.general*) given the distinction made by Freedman & Shah [20] between graphical knowledge and domain knowledge.

Qualitative coding. After reaching inductive thematic saturation on the coding scheme [51], we coded the walkthrough portions of the interviews using Taguette [48]. The qualitative coding was primarily done by the first author, with extensive discussions with the second author on each

[3]The *goal* subcode was later renamed *takeaways* resulting in *purpose.goal* being renamed *purpose.takeaways*.

Table 1: Summary of the visualization courses for the participants in our final analysis.

Course	Subject	Assessment components	Department	Audience	Class Size	Number of participants
C1 (Univ. A)	Interactive information visualization	Assignments, readings, quizzes, final project	Computer science	Upper level undergrads and grad students	∼40	2
C2 (Univ. A)	Data visualization	Assignments, readings, labs, final project	Statistics	Statistics/data science undergrads and grad students (prereq: intro statistics)	∼70	2
C3 (Univ. B)	Data visualization	Assignments, readings, final project	Computer science	Third and fourth years undergrads	∼130	6
C4 (Univ. B)	Information visualization	Assignments, readings, labs, quizzes, final project	HCI	Upper level undergrads	∼40	5
C5 (Univ. C)	Information visualization	Assignments, readings, quizzes, research presentation, final project	Computer science	Fourth year undergrads	∼20	1

part of the process, particularly anything unclear or ambiguous. When coding the walkthroughs, **file names were anonymized to minimize bias** based on knowledge of whether the walkthrough was from a PRE or a POST interview. Because the first author conducted the interviews and transcripts were looked at throughout developing and refining codes, it would be impossible to remove all bias. However, given the similarities in the content of the interviews and the amount of time that had passed between code development and final coding, the coder did not have a conscious recollection of which participant each transcript was from, and whether it was from PRE or POST. The coding scheme and coding were both finalized before de-anonymizing the transcripts and comparing the frequencies of each code from PRE to POST.

4.2 Modified VLAT Analysis

Because of our small sample size, we looked only at descriptive statistics of VLAT scores. We report each person's score for each session and the mean score for each session in Sec. 5.2.

5 Results

5.1 Interview qualitative analysis

There were 13 higher-level codes: *confusion*, *data*, *examples*, *explain*, *generalizing*, *guessing*, *improve*, *metaphor*, *opinion*, *priorKnowledge*, *purpose*, and *unfamiliar*. Of these, 8 had sub-codes, with a total of 35 sub-codes. In addition to the 5 codes with no sub-codes, this resulted in a final list of 38 unique codes. The higher-level codes are outlined in Tab. 2, and the full coding scheme including all sub-codes is included in the supplemental materials. Fig. 4 shows the difference from PRE to POST in the number of walkthroughs containing each higher-level code.

5.1.1 Walkthroughs in both PRE and POST reflected existing models of graph comprehension

There were many similarities between walkthroughs in PRE and POST. The most frequently occurring codes in both PRE and POST were *visual* and *explain*, and participants in both sessions spent the majority of their time describing the visual features of the graph and providing explanations. Both the PRE and POST walkthroughs reflected existing models of graph comprehension, such as Lee et al.'s NOVIS model [34].

The NOVIS model of graph comprehension [34], inspired by Klein et al.'s data-frame theory of sensemaking [30], includes the steps of "encountering visualization", "constructing a frame", "exploring visualization", "questioning the frame", and "floundering on visualization". Although these steps manifested differently in the two sets of interviews, all five steps appeared in both the PRE walkthroughs and the POST walkthroughs. Often, participants started by describing visual features (the *visual* code) as they were encountering and exploring the visualization, and explained their understanding of how the chart worked (the *explain* code) as they were constructing a frame (their initial interpretation of the visualization). Questioning the frame and floundering on the visualization were primarily encapsulated by the *confusion* code and the *guessing* code, which appeared in both sets of walkthroughs, although more frequently in the PRE walkthroughs.

Here we use P15-POST's walkthrough of SCALE-STACK BAR CHART as an illustrative example of our coding scheme and how these codes are reflected by the NOVIS model [34]:

> Age structure of selected countries. [*visual.title*] So population, 0 to 100k, 10 million, 100 million, 500 million, 2 billion, 65 plus. [*visual.axes, visual.scale, visual.labels*] Okay. Selected countries age structures. [*purpose.topic*] Okay. So Canada. I see broken bar charts and a legend on the right. So it gives the age range, which I think is pretty straightforward. [*visual.other, examples.senseOne*] And then the countries on the bottom, what I'm a little confused on is that, right away I'm not able to tell, for example, looking at Canada, this would be empty, this would be empty, which to me it means, oh, maybe no population. [*examples.senseOne, confusion.how*]
>
> And then seeing a chart here, in my head, I would think that maybe Canada's population is around 30 to 40 million people [*confusion.interpretation*], but then, oh, so the chart just shrinks. [*explain.how*] Oh, I see. Yes, I think it's pretty straightforward then in this case, over here is a zoomed in photo and then over here is more zoomed out as we zoom out on the axis. [*explain.how*] And then these colors represent the age makeup of the population. [*explain.what, visual.color*]

P15-POST begins by reading the title (*visual.title*) and axis labels to get a sense of the scale (*visual.axes*; *visual.scale*; *visual.labels*). They transition to looking at one specific example, Canada, to help make sense of the chart (*examples.senseOne*). In NOVIS, this would be "encountering the visualization" and "exploring the visualization".

They notice that Canada's population is empty for some of the boxes, and they express confusion about how to read this, as it makes them think there is no population (*confusion.how*). They express confusion about that interpretation given that another part of the chart indicates that Canada's population is around 30 to 40 million people (*confusion.interpretation*). In NOVIS, this would be "questioning the frame".

Finally, they realize how the chart works, and explain how to read it, saying "over here is a zoomed in photo and then over here is more zoomed out" (*explain.how*). They also explain what the colors represent (*explain.what*). Here, they have constructed a new frame, this time correctly, and they use it to explain how the chart works.

Our findings corroborate existing models, like NOVIS, but also allow us to examine how students' graph comprehension changes as they learn visualization skills—as we describe next.

5.1.2 PRE participants expressed more confusion, especially lower-level confusion

The *confusion* code describes portions of the interviews in which participants describe a source of confusion. PRE participants expressed more confusion overall (*confusion*), especially confusion about what the visuals were showing (*confusion.what*). For example, P1-PRE was not sure

Table 2: The higher-level qualitative codes, a description and example of each, and a list of sub-codes if applicable.

Higher-level code	Description	Sub-codes	Example
Confusion	Expressing confusion or lack of understanding	general, how, interpretation, what, why	"I'm not sure why there are multiple bars for each country" (P1-PRE)
Data	Discussing data	source, type	"It has something to do with nominal variables" (P2-POST)
Examples	Listing specific examples	explainContrasting, explainOne, explainSimilar, senseMultiple, senseOne	"So like for Canada, it's in the hundreds of millions." (P16-PRE)
Explain	Explaining the chart	how, modify, tentative, verify, what, why	"Each bar corresponds to a country below" (P10-PRE)
Generalizing	Generalizing	N/A (no sub-codes)	"And we do the same process for each country" (P15-PRE)
Guessing	Expressing lack of certainty (usually used with another code)	N/A (no sub-codes)	"I'm guessing that might mean the change" (P11-POST)
Improve	Proposing improvements	N/A (no sub-codes)	"They could have different colors for different countries" (P7-POST)
Metaphor	Making visual metaphors	N/A (no sub-codes)	"The bars look like snakes" (P14-POST)
Opinion	Expressing personal opinions	negative, neutral, positive	"Sort of bizarre to be honest. Pretty misleading." (P4-PRE)
PriorKnowledge	Using prior knowledge	general, otherCharts	"I see what looks somewhat similar to a bar graph" (P16-POST)
Purpose	Listing purpose or takeaways	takeaways, topic	"So it's comparing population and age, it looks like." (P11-PRE)
Unfamiliar	Expressing lack of familiarity	N/A (no sub-codes)	"I've actually never seen a graph like that before" (P13-PRE)
Visual	Describing visual features	axes, color, labels, other, salience, scale, structure, title	"I see some dots and then some connected dots" (P2-PRE)

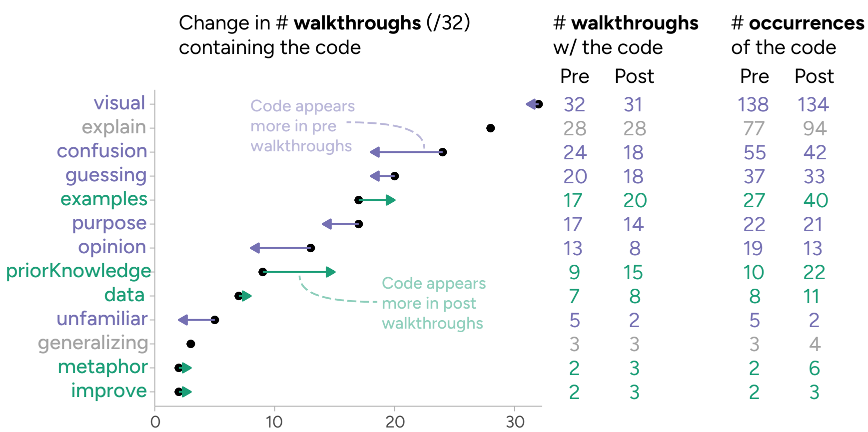

Fig. 4: Change in code frequencies from PRE to POST. The first column shows the change in the number of walkthroughs[4] each code appeared in—an arrow pointing left means the code appeared more in the PRE walkthroughs and vice versa. The second column shows the same data in table form, and the third column shows the raw number of occurrences of each code in PRE and POST.

what the right portion of UPSET represented, saying that they "actually [had] no idea what the whole right part is doing" (*confusion.what*). When POST participants expressed confusion, they tended to express confusion about a particular interpretation of the chart (*confusion.interpretation*). For example, during P11-POST's walkthrough of MARIMEKKO SLOPE CHART, they said "looking at this, it would make me think that there's more people with a bachelor's degree than some college, but I know that doesn't make sense" (*confusion.interpretation*). Here, they focus on the specific source of their confusion, which is that a finding drawn from their initial interpretation seems incorrect. PRE participants tended to express their confusion more broadly and about what the visuals represent, and POST participants were more likely to express confusion about a particular interpretation of the chart, demonstrating an ability to narrow their focus to specific parts of the visualization.

5.1.3 POST participants explained more and went deeper in their explanations

The *explain* code captures participants' explanations of the chart, and Fig. 5 shows the frequencies of *explain* and its subcodes. Both PRE and POST participants used explanations, but POST participants spent more time explaining and were more likely to provide evidence when doing so (*explain.verify*). For example, P10-POST verified their explanation of WRAPPED BAR CHART by saying "I think that checks out based on what I know about COVID cases for each of these countries". POST walkthroughs were also more likely to include an explanation of why a visual choice was made (*explain.why*). For example, P5-POST explained why SCALE-STACK BAR CHART has multiple components, saying "the other countries would just be invisible if they tried to make it scale properly, so they made multiple different stacked plots". Here, P5-POSTxplains why there are multiple axes while also explaining the benefits of this approach over other ways of presenting the data. While PRE and POST participants both explained surface level features of the

[4] Each walkthrough represents one combination of participant and visualization, with four walkthroughs per participant: two in PRE and two in POST.

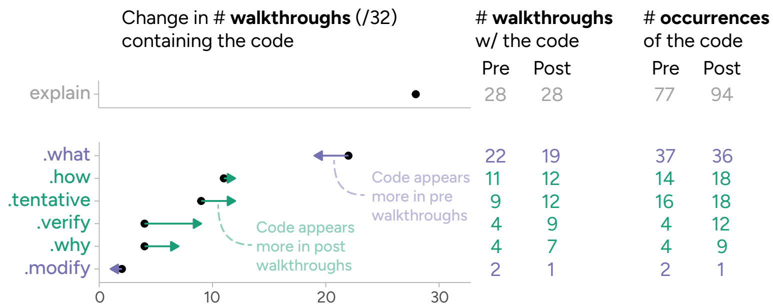

Fig. 5: Change (from PRE to POST) in number of walkthroughs containing *explain* and its subcodes, and total occurrences of each code in each session.

charts, POST participants tended to go deeper and explain why those choices were made. POST participants demonstrated design empathy by thinking about the choices that the chart designers made, often in the context of other possible choices, which demonstrates an understanding of the process of designing an effective visualization.

5.1.4 POST participants used more examples to make sense of and to explain visualizations

The *examples* code was used when participants used examples from within the chart to help make sense of it (e.g. *examples.senseOne*, which captures the use of one example from within the chart for sensemaking) or to explain the chart (e.g. *examples.explainContrasting*, which captures the use of multiple contrasting examples from within the chart to explain it), and POST participants used more examples overall. P9-POST, after expressing confusion about each country in SCALE-STACK BAR CHART having multiple bars (*confusion.what*), focused on Sweden to help make sense of the chart, saying "this one, which I think is Sweden, [...] has multiple up here and I'm not sure how one country can have multiple populations" (*examples.senseOne*). P16-POST, in their explanation of MARIMEKKO SLOPE CHART, first started by stating the goal of the chart, saying that the chart "is showing how the median income for that bucket of people has progressed over the years" (*purpose.topic*). They then enhanced their explanation using a specific example, saying "the left side of the next bar is how much high school graduates were making in 2012, and the right side of the yellow bar is how much they were making in 2016, and so on" (*examples.explainOne*). The use of examples by POST participants, both to make sense of and to explain the visualizations, indicates their tendency to focus more on specific aspects of a chart rather than the chart as a whole, and to be able to deconstruct a visualization into its components.

5.1.5 POST participants talked more about familiar chart types

During the walkthroughs, POST participants talked more frequently about other chart types they were familiar with (*priorKnowledge.otherCharts*) For example, P16-POST began their walkthrough of MARIMEKKO SLOPE CHART by saying "So it's kind of like a histogram, that's what it looks like", drawing on prior knowledge as a starting point for making sense of the chart (*priorKnowledge.otherCharts*). P7-POST compared WRAPPED BAR CHART to a traditional bar chart to explain the benefits of the chart design: "if it were a traditional bar graph, you wouldn't even be able to see Mexico or Egypt or Argentina" (*priorKnowledge.otherCharts*). By explicitly making comparisons to other chart types, participants were able to focus on the unique aspects of the unfamiliar visualization. They also demonstrated their ability to apply their existing knowledge about particular visualization types to a new context—a concept known as *transfer* [36]. Deconstructing a visualization to focus on the less familiar aspects is a strategy that may help in making sense of a visualization, as well as in explaining it.

5.1.6 PRE participants focused on data source and POST participants focused on data type

The *data* code was used when participants discussed the data used in the visualization. As shown in Fig. 6, PRE participants were more likely

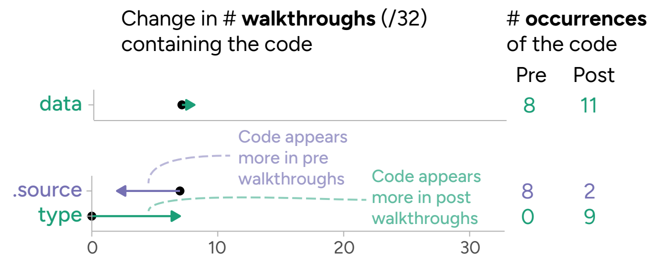

Fig. 6: Change (from PRE to POST) in number of walkthroughs containing *data* and its subcodes, and total occurrences of each code in each session.

to discuss the source of the data (*data.source*), whereas only POST participants talked about the data type (*data.type*). When PRE participants talked about the source of the data (*data.source*), it was usually in the context of understanding how the data were collected. For example, P13-PRE appeared to assume that the data for MARIMEKKO SLOPE CHART were collected using a survey, saying that "anyone [...] over 25 years old responded to this". Similarly, P1-PRE said that UPSET was "representing data [...] from a survey done in Toronto". POST participants only discussed the data source when explaining why a particular visual choice was made. P1-POST, for example, explained why MARIMEKKO SLOPE CHART only included data for people over 25, saying "they just calculated people who are already 25 years old or more so that they don't calculate people who haven't finished their degree yet." (*explain.why*, *data.source*)

Only POST participants mentioned data type. When looking at WRAPPED BAR CHART, P2-POST said that "On the left side, there's some quantitative variables, which makes sense because the title talks about the total number of cases." (*visual.title*, *data.type*, *visual.axes*, *explain.verify*), using what they saw on the axis to determine the data type, and using the title to verify that explanation. Participants used specific terminology to describe the data type. P2-POST, for example, said that MARIMEKKO SLOPE CHART was "mapping to the nominal variable at education level". P10-POST, when looking at WRAPPED BAR CHART, noted that "it's categorical stuff on the X axis". Data types are often explicitly taught in visualization classes, using words like "nominal" and "categorical" (e.g., Munzner's commonly-used textbook employs these terms [42]), and data types were taught in all the courses we recruited from. The fact that participants were thinking about data types suggests both that participants acquired vocabulary specific to data visualizations, but also that they were exhibiting a more sophisticated understanding of the relationship between the visual domain and the data domain, which is a focus of many definitions of visualization literacy [7].

5.1.7 PRE participants expressed more negative opinions

In PRE, participants expressed negative opinions about the chart (*opinion.negative*) slightly more frequently than in POST. For example, in their walkthrough of SCALE-STACK BAR CHART, P4-PRE described the visualization as "annoying", saying that "you can only really compare the countries to each other when they have the same y axis", which they described as "sort of bizarre" and "pretty misleading".

P1-POST, on the other hand, focused their negative opinion on a specific aspect of the visualization, and explained the reasons they felt the design choice was unnecessary. In describing MARIMEKKO SLOPE CHART, they said that "the colors are a bit unnecessary, because they're already labeled clearly and divided into different bars with captions" (*improve*, *opinion.negative*, *visual.color*). Although this code appeared fairly infrequently, the tendency for students to be more negative before taking a visualization class is something that the authors have also anecdotally noticed. It is often easy for new students to nitpick a design; what is harder is to make a careful judgment about what is good and bad about a design, taking into account the constraints the designer was under. This more considered form of judgment is hard to gain except by designing visualizations, as our participants did in their visualization classes.

5.2 VLAT scores

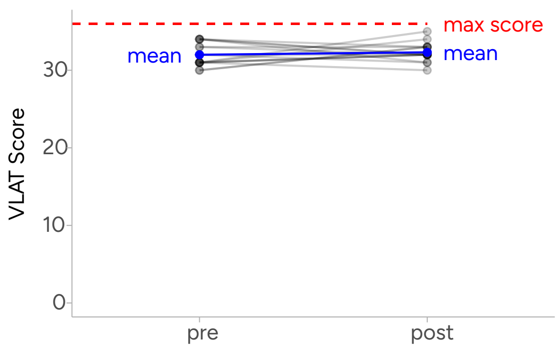

Fig. 7: VLAT scores reflect a ceiling effect; scores are clustered near the maximum score. There is no evidence of improvement from PRE to POST.

Results on the VLAT are shown in Fig. 7. During the PRE session, the mean score was 32 out of 36 points, and during the POST interview, the mean score was 32.31 out of 36 points. The scores ranged from 30 to 34 for PRE participants, and from 30 to 35 for POST participants. This indicates a likely ceiling effect, as initial scores were very high. In the original VLAT paper, raw scores ranged from 14 to 50 out of 54, with a mean score of 34.72 out of 54 [35].[5] Given that our participants averaged around 90% of the questions correct and the population that VLAT was tested on averaged around 65% of the questions correct (including the easier questions we excluded), it is clear that our sample was more skilled at the abilities VLAT assesses than the general public. There is no clear evidence of participants' VLAT scores systematically changing after taking a visualization class, suggesting that their visualization classes may not have helped them improve on the construct being measured by VLAT.

6 IMPLICATIONS

6.1 Students' design empathy and deconstruction ability improved

Although our participants were from a variety of visualization classes, all classes aimed to improve students' ability to work with visualizations, and all classes emphasized identifying what makes a visualization effective, deconstructing a visualization into data types and encodings, visualization design, visualization critique, and visualization implementation in a range of technical tools. Prior work on visualization and statistics education has listed teaching goals such as "visualization principles" (e.g. perception, design principles, critique), "design, programming, and system building skills" [5], "help[ing] students learn how to make good graphs", and "incorporat[ing] the use of graphics in data discovery" [43], all of which are components of the classes we recruited from.

We noticed several qualitative differences between the PRE and POST walkthroughs, indicating that the visualization courses did seem to affect how students reasoned about unfamiliar visualizations. One key difference was that after taking a visualization class, participants were better able to **deconstruct** visualizations, and tended to narrow in on specific details both in making sense of the visualizations and in explaining them. All of the visualization classes taught students about the building blocks of visualizations (e.g. encodings and data types), and two classes used the grammar of graphics to do so [58, 59], explicitly introducing students to the notion of grammar and syntax in visualization. This could explain why students got better at breaking the chart down into its components. The ability to effectively deconstruct visualizations reflected students' budding expertise, as experts are generally better at "chunking" information into familiar patterns [8, 38].

Another key difference after taking a visualization class was that students were better able to empathize with visualization designers when reading visualizations, a skillset we are referring to as **design empathy**. Discussions of data type (*data.type*) in the POST walkthroughs suggest that participants were thinking about how the visualizations were created, as many design decisions are based on the type of data the designer is visualizing. POST participants also empathized with the designer when explaining why the visualization was designed in a particular way (*explain.why*) and acknowledging design tradeoffs rather than criticizing small choices (*opinion.negative*). Both of these key differences indicate that students were engaging with visualizations in more sophisticated ways after taking a visualization class.

6.2 Incorporating deconstruction and design empathy into visualization education

Our findings suggest that students learned to deconstruct visualizations and exhibit design empathy. While attributing our results to specific aspects of the classes was beyond the scope of this study, our findings provide insight into topics that visualization educators may want to emphasize. The courses that explicitly listed learning outcomes on their syllabi included learning outcomes related to more abstract and cognitively demanding processes, such as creating and critiquing visualizations (mirroring the highest levels—'create' and 'evaluate'—of Bloom's Revised Taxonomy [32]). However, our key findings that students are better able to deconstruct visualizations and understand tradeoffs made by visualization designers were not reflected by any of the explicit learning outcomes in the syllabi. Although these topics are likely implicitly addressed in visualization classes, we encourage instructors to consider explicitly adding them as learning outcomes to ensure their continued emphasis in their classes. Instructors could highlight these points when they come up—for example, when talking about a particular visualization, they could explicitly highlight that each of the choices were made by the designer from a set of many possible options. Visualization instructors could also consider activities such as discussing the designer's potential reasoning for each choice in a visualization, or asking students to deconstruct a visualization by using a framework such as the grammar of graphics [58, 59] to describe its components.

Outside of the university context, the skills of deconstructing visualizations and taking the designer's perspective are likely to be helpful in learning to more effectively read visualizations, and we would be interested in seeing how they could be applied in K-12 classes and in shorter workshops for the general public.

Prior work has developed online platforms and games for teaching visualization skills to children—both for reading and constructing visualizations [1, 6, 25]. It would be interesting to see if similar tools could be built, either for children or for adults, that focus on deconstructing a visualization and thinking about the data types and encoding choices within it. Tools like these could be a valuable addition to visualization courses or to visualization interventions for the general public.

6.3 Measuring other aspects of visualization literacy, including deconstruction and design empathy

While visualization instructors have a variety of methods for assessing their students, validated measurement instruments in the style of VLAT have the potential to be valuable for evaluating the quality of teaching methods and determining if students are improving on a standardized scale. However, despite qualitative evidence of participants engaging in visualizations in more sophisticated ways, we did not find evidence

[5]These scores are for the test tryout phase of VLAT test development, as they only report corrected scores and not raw scores for the final 53-item version.

that VLAT scores systematically improved. The scores were clustered closely together, ranging from 30–35 out of 36 points, indicating a likely ceiling effect. It may have been more effective to use a more difficult test, or a test that has been validated for use with university students. Ge et al.'s CALVI assessment, for example, which measures "the ability to read, interpret, and reason about erroneous or potentially misleading visualizations," is more difficult than VLAT, and therefore may not have the same ceiling effect on this population [24].

Given our findings and prior statements about the limitations of VLAT [24, 35], it is likely that VLAT and other similar assessments measure a construct that is different from the skillset students in data visualization classes develop. Visualization classes often focus on what makes charts effective rather than how to read specific chart types, and VLAT does not measure someone's understanding of the effectiveness of a chart. VLAT is based on visualization tasks that are related to "reading and interpreting visually represented data", such as "Retrieve Value", "Determine Range", and "Find Correlations/Trends", and applies those tasks to relatively common visualization types, not complex or novel visualizations like those often studied in university-level visualization classes. VLAT and other existing assessments [4,7,23,24,31,35,39,44] measure only a subset of the previously-listed learning goals for visualization classes, and none of the assessments measure the dimensions on which we found qualitative improvements, such as the ability to deconstruct a chart into its constituent parts and to exhibit design empathy when engaging with visualizations. These gaps may indicate that there are aspects of visualization literacy that are not captured by existing measurement instruments. A systematic survey of current visualization literacy assessments and the competencies they evaluate could make it easier to select appropriate assessments for a given use case and to understand the gaps in existing assessments [26].

As mentioned in Sec. 6.2, visualization courses cover a range of learning outcomes across different levels of understanding; new measurement instruments could mirror this by evaluating these different levels of understanding. While visualization literacy tests such as VLAT primarily focus on skills covered by the "Remember" and "Understand" levels in the Revised Bloom's taxonomy [32], future work could develop other tests to cover the more cognitively complex levels. This may include how well students can critique visualization designs ("Evaluate") or create new visualizations ("Create"). These abilities are often assessed in visualization classes through the completion of assignments and projects, and there would be value to creating validated self-contained tests inspired by these classroom assessments. In addition, future work could create assessments that explicitly measure abilities that we found students learned, such as deconstructing visualizations and taking the perspective of a visualization designer. For example, a test could be designed in which participants see a visualization and have to explicitly justify why the designer made particular choices (e.g. "Why did the designer choose to facet by variable x?"). Many of these abilities are difficult to measure in a multiple-choice format, and another valuable area of future work is to find scalable ways to assess these more cognitively complex visualization abilities.

7 Discussion and limitations

7.1 Broadening our understanding of visualization literacy

We qualitatively found improvements in the ways students discussed graphs, but these differences were not reflected in VLAT scores and were in areas not measured by existing assessments of visualization literacy. Pandey & Ottley say that "visualization literacy is a multidimensional construct, and measuring its full scope with a single scale is untenable" [45]. We agree—the current visualization literacy assessments are useful, but do not offer a complete picture. Even if other aspects of visualization literacy (e.g. visualization critique or deconstruction) are more difficult to assess, we should acknowledge that existing tests measure only a specific component of visualization literacy.

In a discussion of new kinds of literacy, Taylor argues that literacy needs to involve the ability not just to read, but also to understand and construct communication [53]. Statistical literacy emphasizes contextual understanding and critical thinking skills, particularly the use of critical thinking skills in everyday life [21, 52, 57]. These components are all relevant to visualization literacy, and could be incorporated in how we think about it. In Sec. 6.3, we discuss possible directions for assessments of a wider range of visualization skills, and these skills should all be considered part of visualization literacy.

More broadly, Vee describes a literacy as being "necessary for everyday life" [55], and Roth & McGinn talk about graphing as a "shared social practice" [50]. The role of graphing in people's lives is another important component of visualization literacy to consider. This is already reflected in work such as Börner et al.'s study interviewing museum visitors about the ways they have encountered visualizations in their own lives [11]. Peck et al., from interviews with people in rural communities about their attitudes and perceptions of data visualizations and their usefulness, discuss the ways people talk about visualizations, the ways they think critically about the source of the data, and the ways they think about their own relationship to the visualizations [46]. These are examples of ways that people engage with visualizations that should also be reflected in our conceptualization of visualization literacy.

7.2 Limitations

Because of the logistical challenges of a study like ours, there are limitations to our findings. The longitudinal study design means there are inherent differences between the two sessions. The POST sessions, for example, were conducted during the exam period for most participants, which would likely change their level of fatigue and focus. While we are able to counterbalance the effects of chart type and participant, we are not able to counterbalance systematic differences between PRE and POST.

Another limitation is that the four unfamiliar visualizations were fairly different from each other, in terms of both content and ease of comprehension with descriptions removed. Almost all participants correctly understood how WRAPPED BAR CHART worked, whereas the majority had trouble with UPSET and MARIMEKKO SLOPE CHART. These differences made it difficult to come up with a coding scheme that covered all four charts, or to compare interviews of any individual participant in their PRE walkthrough and their POST walkthrough.

Because of the qualitative nature of our work, we had a relatively small number of participants in our study, all from R1 universities, which may limit the generalizability of our findings. Participants' high VLAT scores indicate that they are not representative of the audience that VLAT was designed for, and they are likely also not representative of all university students. Due to privacy concerns, we did not collect data on students' academic performance in the course, which also limits our ability to understand the representativeness of our sample.

Given the difficulty of finding a natural setting in which people improve on visualizations to conduct a longitudinal study, we chose to look at university-level visualization classes, and most of our participants had prior experience in statistics and an interest in data visualization. It would be interesting to see how these results look in a different setting where people with less familiarity with visualizations go through some kind of training about visualizations, perhaps a workshop or tutorial.

We studied multiple visualization classes to look at visualization literacy generally, rather than to evaluate one specific class, which means we cannot draw direct parallels between course content and our findings. Future work could vary specific interventions or activities in a controlled setting to see how they impact the development of specific skills. Bach et al. echo this call to action, expressing a need for "reliable empirical evidence about the effectiveness of our [educational] approaches" [2].

8 Conclusion

Our aim in this paper was to understand how visualization literacy develops in students in university-level visualization classes. We found qualitative changes in how students approached visualizations, with POST participants exhibiting many of the skills that visualization classes aim to teach and that indicate comfort with graphs, such as thinking about the perspective of the designer and deconstructing a visualization into its components. Although current quantitative measures of visualization literacy do not aim to measure skills such as deconstruction and design empathy, and they would in fact be very difficult to measure quantitatively, we suggest that the visualization community would benefit from assessments for these skills.

9 Supplemental Materials

Our supplemental materials are available at `https://osf.io/w5pum/`. We provide our walkthrough protocol, modified VLAT questions, edited and unedited versions of the unfamiliar visualizations, walkthrough transcripts, qualitative codebook, qualitative coding data, VLAT data, and analysis code.

Acknowledgments

The authors wish to thank the instructors of the classes we recruited from, and members of the MU Collective lab for their feedback. This work was supported in part by a grant from the National Science Foundation (#2120750).

References

[1] B. Alper, N. H. Riche, F. Chevalier, J. Boy, and M. Sezgin. Visualization Literacy at Elementary School. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pp. 5485–5497. ACM, Denver Colorado USA, May 2017. doi: 10.1145/3025453.3025877 2, 8

[2] 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 Transactions on Visualization and Computer Graphics*, pp. 1–12, 2023. doi: 10.1109/TVCG.2023.3327378 9

[3] F. B. Baker. *The basics of item response theory*. ERIC, 2001. 3

[4] C. A. Berg and P. Smith. Assessing students' abilities to construct and interpret line graphs: Disparities between multiple-choice and free-response instruments. *Science Education*, 78(6):527–554, Nov. 1994. doi: 10.1002/sce.3730780602 2, 9

[5] J. Beyer, H. Strobelt, M. Oppermann, L. Deslauriers, and H. Pfister. Teaching visualization for large and diverse classes on campus and online. In *Proceedings of IEEE VIS workshop on pedagogy data visualization*, 2016. 2, 8

[6] 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 Transactions on Visualization and Computer Graphics*, 26(1):451–460, Jan. 2020. doi: 10.1109/TVCG.2019.2934804 8

[7] J. Boy, R. A. Rensink, E. Bertini, and J.-D. Fekete. A Principled Way of Assessing Visualization Literacy. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1963–1972, Dec. 2014. doi: 10.1109/TVCG.2014.2346984 1, 7, 9

[8] J. D. Bransford, A. L. Brown, and R. R. Cocking. How experts differ from novices. *How people learn: Brain, mind, experience, and school*, pp. 31–50, 2000. 8

[9] A. Burns, C. Xiong, S. Franconeri, A. Cairo, and N. Mahyar. How to evaluate data visualizations across different levels of understanding, Sept. 2020. 2

[10] 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, Feb. 2019. doi: 10.1073/pnas.1807180116 1, 2

[11] 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, July 2016. doi: 10.1177/1473871615594652 2, 9

[12] P. A. Carpenter and P. Shah. A model of the perceptual and conceptual processes in graph comprehension. *Journal of Experimental Psychology: Applied*, 4(2):75, 1998. 2

[13] 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. 1

[14] F. Chevalier, N. H. Riche, B. Alper, C. Plaisant, J. Boy, and N. Elmqvist. Observations and reflections on visualization literacy in elementary school. *IEEE computer graphics and applications*, 38(3):21–29, 2018. 2

[15] J. R. Conway, A. Lex, and N. Gehlenborg. UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics*, 33(18):2938–2940, 06 2017. doi: 10.1093/bioinformatics/btx364 3

[16] Y. Cui, L. W. Ge, Y. Ding, F. Yang, L. Harrison, and M. Kay. Adaptive Assessment of Visualization Literacy. *IEEE Transactions on Visualization and Computer Graphics*, 30(1):628–637, Jan. 2024. doi: 10.1109/TVCG.2023.3327165 4

[17] L. Ellis. Set analysis: A face off between venn diagrams and upset plots, Apr 2019. 3

[18] J. Fereday and E. Muir-Cochrane. Demonstrating rigor using thematic analysis: A hybrid approach of inductive and deductive coding and theme development. *International journal of qualitative methods*, 5(1):80–92, 2006. 4

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

[20] E. G. Freedman and P. Shah. Toward a model of knowledge-based graph comprehension. In *International conference on theory and application of diagrams*, pp. 18–30. Springer, 2002. 2, 4

[21] I. Gal. Statistical literacy: Meanings, components, responsibilities. In *The challenge of developing statistical literacy, reasoning and thinking*, pp. 47–78. Springer, 2004. 9

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

[23] R. Garcia-Retamero, E. T. Cokely, S. Ghazal, and A. Joeris. Measuring Graph Literacy without a Test: A Brief Subjective Assessment. *Medical Decision Making*, 36(7):854–867, Oct. 2016. doi: 10.1177/0272989X16655334 2, 9

[24] L. W. Ge, Y. Cui, and M. Kay. CALVI: Critical Thinking Assessment for Literacy in Visualizations. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pp. 1–18. ACM, Hamburg Germany, Apr. 2023. doi: 10.1145/3544548.3581406 2, 9

[25] 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 *Extended Abstracts of the Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts*, pp. 389–396. ACM, Barcelona Spain, Oct. 2019. doi: 10.1145/3341215.3356283 8

[26] M. Hedayati, A. Hunt, and M. Kay. From pixels to practices: Reconceptualizing visualization literacy, May 2024. doi: 10.31219/osf.io/6mq42 1, 9

[27] M. Hlawatsch, F. Sadlo, M. Burch, and D. Weiskopf. Scale-stack bar charts. *Computer Graphics Forum*, 32(3):181–190, 2013. doi: 10.1111/cgf.12105 3

[28] A. Karduni. Visualizing categorical data with disproportionate values using du bois wrapped bar charts, Apr 2020. 3

[29] A. Karduni, R. Wesslen, I. Cho, and W. Dou. Du Bois wrapped bar chart: Visualizing categorical data with disproportionate values. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, 12 pages, p. 1–12. Association for Computing Machinery, New York, NY, USA, 2020. doi: 10.1145/3313831.3376365 3

[30] G. Klein, J. Phillips, E. Rall, and D. Peluso. A data-frame theory of sensemaking. *Expertise out of Context: Proceedings of the Sixth International Conference on Naturalistic Decision Making*, pp. 113–155, Jan. 2007. 5

[31] B. Kramarski and Z. R. Mevarech. Enhancing Mathematical Reasoning in the Classroom: The Effects of Cooperative Learning and Metacognitive Training. *American Educational Research Journal*, 40(1):281–310, Jan. 2003. doi: 10.3102/00028312040001281 2, 9

[32] D. R. Krathwohl. A Revision of Bloom's Taxonomy: An Overview. *Theory Into Practice*, 41(4):212–218, 2002. 8, 9

[33] M. Lambrechts, Oct 2018. 3

[34] S. Lee, S.-H. Kim, Y.-H. Hung, H. Lam, Y.-A. Kang, and J. S. Yi. How do People Make Sense of Unfamiliar Visualizations?: A Grounded Model of Novice's Information Visualization Sensemaking. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):499–508, Jan. 2016. doi: 10.1109/TVCG.2015.2467195 2, 4, 5

[35] S. Lee, S.-H. Kim, and B. C. Kwon. VLAT: Development of a visualization literacy assessment test. *IEEE transactions on visualization and computer graphics*, 23(1):551–560, 2017. 1, 2, 3, 8, 9

[36] S. Lewandowsky, D. Little, and M. L. Kalish. Knowledge and expertise. *Handbook of applied cognition*, pp. 83–109, 2007. 7

[37] A. Lex, N. Gehlenborg, H. Strobelt, R. Vuillemot, and H. Pfister. Upset: Visualization of intersecting sets. *IEEE Transactions on Visualization and Computer Graphics (InfoVis)*, 20(12):1983–1992, 2014. doi: 10.1109/TVCG.2014.2346248 3

[38] R. Lowe. "Reading" scientific diagrams: Characterising components of skilled performance. *Research in Science Education*, 18(1):112–122, 1988. 8

[39] A. V. Maltese, J. A. Harsh, and D. Svetina. Data Visualization Literacy: Investigating Data Interpretation Along the Novice—Expert Continuum. *Journal of College Science Teaching*, 45(1):84–90, 2015. 2, 9

[40] A. McCann. Marimekko slope chart, Aug 2018. 3

[41] B. Miles Matthew, H. A. Michael, and S. Johnny. Qualitative data analysis: A methods sourcebook, 2014. 4

[42] T. Munzner. *Visualization analysis and design*. CRC press, 2014. 3, 7

[43] D. Nolan and J. Perrett. Teaching and Learning Data Visualization: Ideas and Assignments. *The American Statistician*, 70(3):260–269, July 2016. doi: 10.1080/00031305.2015.1123651 2, 8

[44] A. V. Pandey, K. Rall, M. L. Satterthwaite, O. Nov, and E. Bertini. How Deceptive are Deceptive Visualizations?: An Empirical Analysis of Common Distortion Techniques. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pp. 1469–1478. ACM, Seoul Republic of Korea, Apr. 2015. doi: 10.1145/2702123.2702608 2, 9

[45] S. Pandey and A. Ottley. Mini-VLAT: A Short and Effective Measure of Visualization Literacy. *Computer Graphics Forum*, 42(3):1–11, June 2023. doi: 10.1111/cgf.14809 4, 9

[46] E. M. Peck, S. E. Ayuso, and O. El-Etr. 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*, pp. 1–12. ACM, Glasgow Scotland Uk, May 2019. doi: 10.1145/3290605.3300474 9

[47] S. Pinker. A theory of graph comprehension. *Artificial intelligence and the future of testing*, pp. 73–126, 1990. 2

[48] R. Rampin and V. Rampin. Taguette: open-source qualitative data analysis. *Journal of Open Source Software*, 6(68):3522, 2021. doi: 10.21105/joss.03522 4

[49] W.-M. Roth and G. M. Bowen. Professionals Read Graphs: A Semiotic Analysis. *Journal for Research in Mathematics Education*, 32(2):159, Mar. 2001. doi: 10.2307/749672 4

[50] W.-M. Roth and M. K. McGinn. Graphing: Cognitive ability or practice? *Science Education*, 81(1):91–106, Jan. 1997. doi: 10.1002/(SICI)1098-237X(199701)81:1<91::AID-SCE5>3.0.CO;2-X 9

[51] B. Saunders, J. Sim, T. Kingstone, S. Baker, J. Waterfield, B. Bartlam, H. Burroughs, and C. Jinks. Saturation in qualitative research: exploring its conceptualization and operationalization. *Quality & quantity*, 52:1893–1907, 2018. 4

[52] S. Sharma. Definitions and models of statistical literacy: a literature review. *Open Review of Educational Research*, 4(1):118–133, Jan. 2017. doi: 10.1080/23265507.2017.1354313 9

[53] C. Taylor. New kinds of literacy, and the world of visual information. *Literacy*, 2003. 1, 9

[54] S. B. Trickett and J. G. Trafton. Toward a Comprehensive Model of Graph Comprehension: Making the Case for Spatial Cognition. In D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J. C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, M. Y. Vardi, G. Weikum, D. Barker-Plummer, R. Cox, and N. Swoboda, eds., *Diagrammatic Representation and Inference*, vol. 4045, pp. 286–300. Springer Berlin Heidelberg, Berlin, Heidelberg, 2006. 2

[55] A. Vee. Understanding computer programming as a literacy. *Literacy in Composition Studies*, 1(2):42–64, 2013. 9

[56] H. Wainer. A Test of Graphicacy in Children. *Applied Psychological Measurement*, 4(3):331–340, July 1980. doi: 10.1177/014662168000400305 2

[57] K. K. Wallman. Enhancing Statistical Literacy: Enriching Our Society. *Journal of the American Statistical Association*, 88(421):1–8, Mar. 1993. doi: 10.1080/01621459.1993.10594283 9

[58] H. Wickham. A Layered Grammar of Graphics. *Journal of Computational and Graphical Statistics*, 19(1):3–28, Jan. 2010. doi: 10.1198/jcgs.2009.07098 8

[59] L. Wilkinson. *The grammar of graphics*. Springer, 2 ed., 2005. 8