MASTERARBEIT | MASTER'S THESIS

Titel | Title

Designing and Testing a Multidimensional Assessment of Visual Data Literacy (MAVIL)

verfasst von | submitted by

Antonia Saske BA

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Master of Arts (MA)

Wien | Vienna, 2025

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 066 647
Studienrichtung lt. Studienblatt | Degree programme as it appears on the student record sheet:	Masterstudium Digital Humanities
Betreut von | Supervisor:	Univ.-Prof. Torsten Möller PhD
Mitbetreut von | Co-Supervisor:	Laura Koesten MSc Ph.D.

Abstract

Understanding data visualizations, such as bar charts, can be seen as a crucial ability in current data-driven information culture. This ability, referred to as Visual Data Literacy (VDL), involves reading, interpreting, and critically assessing visualized information. Traditional VDL assessments often rely on task-based performance measures, such as retrieving values from a graph. While useful for measuring specific skills, these methods neglect other ability dimensions well-established in visualization research. These include graph familiarity, self-perceived comprehension, and the interpretation of visualization design – all of which may shape how readers interact with or understand data visualizations. Therefore, MAVIL, a Multidimensional Assessment of Visual Data Literacy, is introduced and build on conventional performance-based measures, but incorporates self-perceived ability ratings across six ability dimensions: *Comprehending*, *Decoding*, *Aestheticizing*, *Criticizing*, *Reading*, and *Contextualizing*. These ability dimensions stem from deconstructing VDL as a process of understanding, referencing frameworks from the learning sciences. MAVIL was designed for general audiences and tested through two phases of survey pilots with eight participants. It was then deployed in a survey representative Austria's age groups from 18 to 74 years and their male-female gender distribution and in a follow-up implementation in the *Digitize!* social science survey with 2373 respondents. This thesis documents the development of MAVIL and the evaluation of its test implementations. The collected data reflect the surveyed population's VDL – combining self-perceived ability ratings, task performance, and open-ended visualization critique – and shows the perception of two climate data visualizations, a line and bar chart. Results reveal that roughly one in four have deficits in comprehending simple data units, about one in five feel unfamiliar with the chart types shown, and self-perceived numeracy significantly relates to data reading performance ($p = 0.0004$).

Kurzfassung

Das Verständnis von Datenvisualisierungen, wie Balkendiagrammen, ist eine zentrale Fähigkeit in der gegenwärtigen, datengestützten Informationskultur. Diese Fähigkeit, die sich als *Visual Data Literacy* (VDL) bezeichnen lässt, umfasst das Lesen, Interpretieren und kritische Bewerten von visualisierten Informationen. Traditionelle Methoden zur Erfassung von VDL beruhen meist auf leistungsorienterten Maßen, etwa dem Ablesen spezifischer Werte aus Diagrammen. Diese Ansätze messen zwar gezielt bestimmte Fähigkeiten, vernachlässigen jedoch andere Dimensionen, die in der Visualisierungsforschung als relevant gelten. Dazu zählen die Vertrautheit mit Diagrammtypen, die selbstwahrgenommene Verständlichkeit und die Interpretation gestalterischer Elemente – all diese Aspekte können entscheidend beeinflussen, wie Leser*innen mit Datenvisualisierungen interagieren oder sie verstehen. Um diese Lücke zu schließen, wird in dieser Arbeit MAVIL („Multidimensional Assessment of Visual Data Literacy") vorgestellt. MAVIL erweitert leistungsbasierte Ansätze durch Selbsteinschätzungen entlang von sechs Fähigkeitsdimensionen: *Verstehen* (*Comprehending*), *Dekodieren* (*Decoding*), *Ästhetik Wahrnehmen* (*Aestheticizing*), *Kritisieren* (*Criticizing*), *Lesen* (*Reading*) und *Kontextualisieren* (*Contextualizing*). Diese Dimensionen basieren auf einer Dekonstruktion von VDL als Verstehensprozess, basierend auf Konzepten aus den Lernwissenschaften. MAVIL wurde in zwei Phasen mit insgesamt acht Teilnehmer*innen pilotiert. Anschließend wurde es in einer Umfrage mit 438 Personen getestet, die die Altersverteilung von 18 bis 74 Jahren sowie die Geschlechterverteilung (männlich/weiblich) der österreichischen Bevölkerung innerhalb dieser Altersgruppen repräsentieren. Ein reduziertes Fragenmodul von MAVIL wurde zudem in der sozialwissenschaftlichen Umfrage *Digitize!* implementiert und von 2373 Personen beantwortet. Die erhobenen Daten spiegeln die VDL der befragten Bevölkerung wider – durch eine Kombination aus Selbsteinschätzungen, leistungsbasierter Daten und offenen Bewertungen von Visualisierungen – und bilden die Wahrnehmung zweier Klimadatenvisualisierungen ab, eines Linien- und eines Balkendiagramms. Die Ergebnisse zeigen unter anderem, dass etwa ein Viertel angbibt, Schwierigkeiten beim Verstehen einfacher Datenangaben zu haben, sich jede fünfte Person mit den dargestellten Diagrammtypen nicht vertraut fühlt und die selbstwahrgenommene numerische Kompetenz signifikant mit der Lesegenauigkeit korreliert ($p = 0{,}0004$).

Contents

Abstract i

Kurzfassung iii

1. Introduction 1

2. Background 5

2.1. Visual Data Literacy as a Multidimensional Ability 5
2.1.1. Visual Data Literacy Ability Facets 6
2.1.2. Visual Data Literacy from a Bipartite Perspective 6
2.1.3. Deconstructing Visaul Data Literacy into Ability Facets 6
2.2. Limitations of Prior Literacy Assessments 8
2.3. Self-Perception in Ability Assessments 9
2.4. Climate Data Visualizations as Assessment Objects 9
2.5. Merging Ability Dimensions of Visual Data Literacy in MAVIL 10
2.5.1. Dimension 1: Comprehending 11
2.5.2. Dimension 2: Decoding 11
2.5.3. Dimension 3: Aestheticizing 12
2.5.4. Dimension 4: Critiquing 13
2.5.5. Dimension 5: Reading 13
2.5.6. Dimension 6: Contextualizing 13

3. Methodology 17

3.1. MAVIL in Practice: Design of a Survey 17
3.1.1. Survey Tasks 18
3.1.2. Selection and Embedding of Climate Data Visualization Examples 19
3.2. Testing I: MAVIL Survey Pilots 20
3.2.1. Purpose and Procedure of Survey Pilots 22
3.2.2. Analysis of Interview Data from Survey Pilots and Learning Excerpts 22
3.3. Testing II: Full MAVIL Application in a Survey 24
3.3.1. Data Collection in Cooperation With a Market Research Organization 24
3.3.2. Quality Criteria for Valid Survey Cases 24
3.3.3. Demographics of the Survey 25
3.4. Testing III: MAVIL Module in the *Digitize!* Survey 25
3.4.1. Survey Collaboration within the *Digitize!* Project 26
3.4.2. Creating a MAVIL Module Version for the *Digitize!* Survey 26
3.4.3. Survey Data Description for the MAVIL Module 27

4. Findings **29**
4.1. Insights from the Full MAVIL Survey 29
4.1.1. Insights into Performance Measures 30
4.1.2. Respondents' Self-Perceived Abilities and Perception of the Surveyed Charts 31
4.1.3. Correlative Exploration of MAVIL Variables 31
4.1.4. Significance of Age and Numeracy 34
4.1.5. Report on Visualization Criticism in the Text Inputs 34
4.2. Insights from the MAVIL Module in the *Digitize!* Survey 36
4.2.1. Contextualizing: Focus on Personal Attitude to Climate Change . 36
4.2.2. Reading, Comprehending, Decoding and Aestheticizing 37

5. Discussion and Conclusion **39**
5.1. Survey-Based Insights from MAVIL 39
5.2. Limitations 40
5.3. Future Work 41
5.4. Conclusion 44

Bibliography **47**

A. Appendix **57**
MAVIL Catalog of Questions (Full Survey, Translation) 57
MAVIL Fragenkatalog (Full Survey, Original) 65

1. Introduction

Understanding data visualizations is a critical skill in today's information culture, as they are frequently used to communicate information and to engage the public [1, 2, 3]. These visualizations represent data in a wide variety of formats and contexts, ranging from bar graphs to scatterplots, crisis visualizations to health data trackers, and simple static plots to complex interactive graphics. Despite their widespread use and importance for effectively communicating critical information [4, 5, 6], there remains a lack of assessment structures to evaluate their understandability and whether readers – particularly non-experts [7] – can interpret them effectively. Existing assessments for Visual Data Literacy [8, 9, 10] often oversimplify this ability by focusing on task-based measurements like value retrieval, failing to account for other well-established facets in visualization research. This work addresses these gaps by developing a multidimensional evaluation framework that examines how public audiences engage with and critique visualized information, with a particular focus on climate data visualizations.

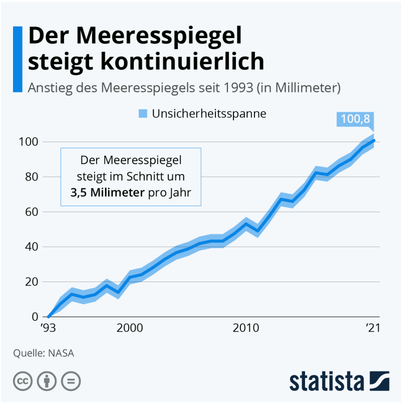
Visual Data Literacy can be defined as the human ability to read, interpret, and critique visual data representations [8]. While research has provided valuable insights into the dimensions of this ability – such as the role of graph familiarity [11] or perceived trust [12] for understanding data visualizations – these insights have not been sufficiently incorporated into existing ability assessments. A common approach to assessing Visual Data Literacy relies on task-based methods, where readers are asked to retrieve specific values, compare data points, or complete similar tasks across multiple visualizations. Their performance is then scored to evaluate their literacy. Examples of such frameworks include the Visualization Literacy Assessment Test (VLAT) [8], its shorter version, the Mini-VLAT [9], and the Critical Thinking Assessment for Literacy in Visualizations (CALVI) [10]. MAVIL complements existing approaches by incorporating self-perceived abilities, enabling an assessment of Visual Data Literacy that considers both task-based performance and how visualization readers perceive their own skills. By incorporating self-perception, MAVIL captures nuances that task-based tests may miss, such as how trustworthy a visualization seems or difficulties linked to specific graph elements.

For the development of MAVIL, Visual Data Literacy is conceptualized as a process of understanding, rooted in the interaction between a visualization reader and a data visualization. Drawing from the learning sciences, this process can be deconstructed into ability facets [13, 14], and each of these dimensions can be turned into a distinct assessment block. The resulting six Visual Data Literacy dimensions covered in MAVIL are as follows:

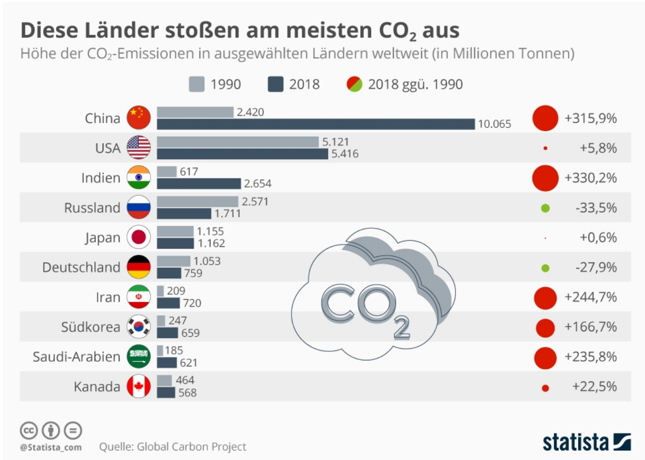
- *Comprehending* covers overall perceived understanding of the visualization and the perceived applicability of its information.
- *Decoding* focuses on the ability to analyze specific visualization components, such as color-encoded graph elements.
- *Aestheticizing* evaluates the respondent's perception of the visualization's aesthetic appeal and its perceived trustworthiness.
- *Critiquing* assesses the ability to critically evaluate the visualization, including articulating suggestions for improvement.
- *Reading* measures the ability to extract and correctly interpret specific data points directly from the visualization.
- *Contextualizing* examines personal interest, background knowledge, and attitudes related to the visualization's content, including broader contextual understanding.

A key area where data visualizations have gained importance in both research and public discourse is climate change and other crises [15, 16, 17]. Visuals can clarify crisis origins, support decision-making, and foster engagement [18, 19, 20], yet balancing scientific accuracy with accessibility remains a challenge [21, 22]. Therefore, climate data visualizations were selected developing and testing MAVIL, as they provide the necessary – and also a critical type of – visual data representation for a literacy assessment. Further, though data visualizations are recognized as meaningful for information communication, there remains a limited understanding of how general audiences engage with them and extract essential information [23, 24]. Prior research on data visualization understanding has focused heavily on expert audiences [11, 22, 7] or specific groups, like students [25, 26]. Given the increasing need for data literacy, as not everyone is in the same way data-savvy [27], our assessment is designed to engage a broad audience, including people of varying ages and without expertise in data visualization. MAVIL is based on implementing different question types (scale ratings for self-perceived abilities, true-or-false questions for data reading tasks, and optional text fields for open-ended responses) to cater to different response styles [28].

The development process of MAVIL covered a design phase and three test implementation phases, all of which were carried out in German language. In the **first test implementation phase**, pilots were conducted with eight participants, who completed the assessment, and discussed their experience in semi-structured interviews. The pilots were conducted to identify potential issues and enhance the assessment's validity through qualitative feedback. They also informed the following **second test implementation phase**, where MAVIL was presented to 438 participants in an online survey. Each participant read data from two climate data visualizations (shown in Figures 1.1a and 1.1b), chosen for their simplicity, familiarity and importance in conveying critical information [15, 19, 29]. This implementation has a representative scope for Austria regarding age



(a) Tested Data Visualization 1: A **line chart** showing the change in sea level between 1993 and 2021, including an uncertainty band.



(b) Tested Data Visualization 2: A **bar chart** comparing CO_2 submissions between 1990 and 2018 for several selected countries.

Figure 1.1.: Summaries of the content and visual features for both Figure 1.1a and Section 1.1b, provided in English, can be found in Section 2.4, while the sources of the tested data visualizations are listed in Section 5.4.

groups (18-74 years) and the male-female gender split within them. In a **third test implementation phase**, parts of MAVIL were included in an interdisciplinary survey. As part of the project *Digitize! Computational Social Sciences in the Digital and Social Transformation*, selected MAVIL questions were presented to 2373 survey participants in Austria. One of the climate data visualizations from the prior implementation, a bar chart (shown in Figure 1.1b), was presented to the survey panel.

While the second and third test implementation phases, where the assessment was tested with survey groups, informed the design of MAVIL, they also provide data for quantitative evaluation. The survey evaluation demonstrates how MAVIL can capture the self-perceived and performance-based ability composition of visualizations readers and it also highlights how participants perceived the tested visualizations, a line chart and bar chart on climate data.

Outlook on Results. In the second test implementation (n=438, ages 18-74), 48% of respondents made mistakes with the simple line and bar charts. About a quarter of respondents found simple data units, such as millimeters, in the surveyed data visualizations incomprehensible. Further, about 12% felt unfamiliar with both displayed graph types. Regarding aesthetic perception, there was a tendency to find the visualizations attractive or well-designed while simultaneously rating them as trustworthy. In the optional text fields on visualization criticism 37% of survey respondents reported at least one point of critique. The qualitative evaluation of text inputs revealed not only respondents' critical abilities but also which visualization elements can cause irritation and how the presented data may evoke emotional responses. In this survey, respondents' self-perceived numeracy levels were significantly related ($p=0.004$) to data reading performance and all other queried dimensions of Visual Data Literacy.

In the third test implementation, where two question blocks and selected self-assessment questions from MAVIL were included in the *Digitize!* social science survey with 2373 respondents, the majority of respondents (69.9%) in Austria agreed that climate change is partly caused by human activity. For around 40% of respondents, reading the climate data from the bar chart was challenging, as they made at least one mistake in three data reading tasks.

Contributions. The design of MAVIL integrates multiple dimensions of the ability to read, interpret, and critique data visualizations into a concise assessment which covers self-perceived and performance-based ability ratings. Its question blocks synthesize findings from various studies and theories on Visual Data Literacy, structuring them into a multidimensional measurement approach. Building on task-based assessments like VLAT [8] and CALVI [10], MAVIL extends beyond data visualization tasks by incorporating self-assessment questions, capturing additional ability facets. The assessment was developed iteratively, integrating qualitative feedback and aiming to be accessible to general audiences. It underwent three test implementation phases that informed the design and resulted in the quantitative analysis of two survey datasets, which are insightful for the lay understanding of two climate data visualizations. These datasets offer insights into the Visual Data Literacy of the surveyed populations and contribute to the broader effort to develop process-oriented assessments of understanding in visualization research.

2. Background

Visual Data Literacy (VDL) has been described as being able to read, interpret and critically deal with data visualizations [8], and many studies and definitions have dealt with this ability (see Section 2.1). For an assessment of VDL, we developed ability dimensions in a theory-driven approach, following both a bipartite framework and multi-faceted descriptions of *understanding* from the learning sciences. The ability assessment through MAVIL, a Multidimensional Assessment of VDL, consists of different question blocks, whereas each question block covers an ability dimension, as outlined in Section 2.5.

Creating an assessment that accounts for the multi-dimensional composition of VDL was deemed necessary because current assessments are dominantly task-based, and therefore, they are perspectively limited in their ability display (see Section 2.2). We highlight studies using self-perceived abilities in relation to performance in Section 2.3, arguing that such ratings can offer unique insights in VDL assessments beyond established methods. When assessing VDL, climate data visualizations were selected as relevant assessment objects because understanding them is a critical scenario in public communication (see Section 2.4).

2.1. Visual Data Literacy as a Multidimensional Ability

Visual literacy has been discussed as an intellectual and social phenomenon since the 1960s, linked to the Information Age and its impact on society and technological culture [30]. Initially, it was explored in education, communication, visual arts, and psychology [30]. These multidisciplinary roots highlight the value of approaching the ability from different perspectives. A multidimensional assessment aligns with this view by treating literacy as an ability that can be described from various perspectives. An exemplary definition of visual literacy in data visualization research from 2017 refers to it as the ability to interpret, understand, and create visual data representations [8]. This aligns with the definition published in the *ACRL Visual Literacy Competency Standards for Higher Education* in 2011, where a visually literate person would be able to "interpret and analyze the meanings of images and visual media" [31] and to "design and create meaningful images and visual media" [31].

In this thesis, we rely on works referencing visual literacy, as this term is well-established, or on those linking literacy to data visualizations (such as [32]). Henceforth, the ability will be referred to as Visual Data Literacy (VDL). To assess VDL, we adopt a bipartite framework, that covers top-down (representing the visualization reader) and bottom-up (representing the visualization) aspects. Additionally, we draw on insights from the learning sciences, which break complex processes, such as understanding, down into sub-skills.

2.1.1. Visual Data Literacy Ability Facets

In order to define ability dimensions of VDL, a literature review was conducted to identify which facets effectively describe the ability, aiming to integrate different theories and study findings. Therefore, VDL is not only defined as the ability to comprehend and interpret data visualizations, but it is influenced by factors such as the complexity of visualized information [22], and abstract thinking skills [33]. Further, the ability encompasses understanding structural elements like data units, visual aspects like color usage [34], and familiarity with graph types [11]. VDL could be related to aesthetic perception, and may include the visual appeal of data visualizations [35] and its potential influence on perceived trustworthiness [12]. Additionally, higher numeracy skills are linked to more accurate data interpretation [36, 37], and personal interest in the displayed topic may also influence the ability to read visual representations [11]. VDL also encompasses the ability to critically evaluate visualizations, including identifying errors and suggesting improvements [38, 39]. All of these ability dimensions may impact a visualization reader's ability to accurately read visual data. For MAVIL, each identified ability dimension was assessed in a question block. There are six dimensions, informed by literature findings, which are made more elaborate in Section 2.5: *Comprehending*, *Decoding*, *Aestheticizing*, *Critiquing*, *Reading*, and *Contextualizing*.

2.1.2. Visual Data Literacy from a Bipartite Perspective

Visual Literacy can be viewed through a bipartite perspective, distinguishing between top-down and bottom-up aspects [40]. Top-down characteristics refer to the perception by visualization readers, shaped by factors such as prior beliefs, demographic characteristics, and personal skills. Here, numeracy, for instance, has been identified as a critical skill influencing data reading accuracy and knowledge gains [37]. In contrast, bottom-up characteristics focus on the visualization itself, including its content or aesthetic structure [34]. MAVIL incorporates both aspects by aligning ability facets with either viewer- or visualization-focused dimensions, as shown in Figure 2.1. For example, Dimension 6, which addresses *Contextualizing* (see the description of MAVIL dimensions in Section 2.5), represents top-down characteristics, and Dimension 2 on *Decoding* represents bottom-up characteristics. We further expanded this two-level approach by integrating frameworks that treat *understanding* as a multifaceted construct in the next Section 2.1.3.

2.1.3. Deconstructing Visaul Data Literacy into Ability Facets

While this bipartite perspective informed MAVIL's design, by addressing both the visualization reader's features and elements of the data visualization, frameworks with a more multi-faceted approach to defining abilities were also integrated. Insights from learning sciences, such as the framework *The Six Facets of Understanding* [13], deconstruct the

Dimension in MAVIL	**Angle**	**Ability Examples Related to Visual Data Literacy**	**Facet of Understanding [13]**	**Example Finding From Full Survey**
Comprehending	Top-Down	Summarize visualized content, recall information [33, 22]	Explanation	76% feel confident in summarizing
Decoding	Bottom-Up	Examine color usage, analyze graph elements [34, 11]	Interpretation	12% feel unfamiliar with both line and bar charts
Aestheticizing	Bottom-Up	Decide upon trustworthiness, judge visual appeal [12, 35]	Empathy	Perceived beauty was linked to trust
Critiquing	Bottom-Up	Articulate criticism, make design suggestions [39, 38]	Perspective	37% submitted meaningful critique in optional inputs
Reading	Top-Down	Solve data reading tasks [8, 9]	Interpretation	48% made errors in data reading tasks
Contextualizing	Top-Down	Utilize statistical skill, put visualization in context [37, 41]	Application	Self-perceived numeracy correlated with data reading

Figure 2.1.: VDL dimensions in MAVIL, developed from either a viewer-focused (top-down) or visualization-focused (bottom-up) perspective. Each dimension is illustrated by example activities and aligned with a corresponding Facet of Understanding [13]. Additionally, we share insights from our representative survey where we tested MAVIL, which is evaluated in detail in Section 4.

complex process of *understanding* by breaking it down into different ability categories. This framework by Wiggins and McTighe conceptualizes *understanding* as being able to explain, interpret, apply, empathize, have perspective, and self-knowledge. Accordingly, a respondent who can summarize the content of a data visualization (see MAVIL Dim1, for an overview of VD ability dimensions refer to Section 2.5) would be able to explain. If a respondent solves a data reading task (Dim5), this is an act of interpretation. Someone who has decided upon the trustworthiness of a visualization (Dim3) has empathized with the perspective of whoever created it. Someone who can contextualize displayed data (Dim1) can apply. Articulating criticism of a visualization, such as making a design change suggestion (Dim4), would be a sign of being able to take perspective. Self-perceived numeracy or interest in climate change means applying existent skills or knowledge (Dim6). Further, some MAVIL dimensions use self-perception ratings to let readers estimate their own ability, which is an excericse that fosters self-knowledge (Dim1, Dim2, Dim6).

Similarly, *Bloom's Taxonomy of Educational Objectives for Knowledge-Based Goals* [14] deconstructs knowledge and learning into ability levels across cognitive, affective, and psychomotor domains, which are most commonly used for cognitive intents [41]. According to this framework, memorizing is followed by understanding, application, then analysis, evaluation, and creation. If applied to VDL, a respondent who recalls the information of a visualization (Dim1) remembers it. Someone who can translate the displayed information by integrating it into their prior knowledge (Dim1) can understand. Solving a data reading task (Dim5) means applying and using information in a new situation. Someone who examines the usage of color (Dim2) can analyze by distinguishing between visualization elements. Someone who justifies a stand, such as making a point of critique (Dim4), can evaluate. A respondent who makes a constructive design change suggestion (Dim4) would be able to create and develop a visualization further.

Following these approaches from the learning sciences, we built MAVIL dimensions by linking facets of understanding to well-established VDL ability examples, as exemplified in Figure 2.1.

2.2. Limitations of Prior Literacy Assessments

Prior literacy assessments on data visualizations often rely on a task-based approach, query a single data visualization type, and/or address specific audience groups. This is the case for recent assessments, such as a treemap literacy test [42] and a parallel coordinate plot literacy test [43]. Several studies focused on an educational context, such as the qualitatively derived principles for a visual literacy assessment in school settings [25]. In another study, it was inquired how expert audiences assessed their understanding of visualized scientific climate change information [11]. Though such attempts were made, they implemented similar, one-dimensional assessment questions. And, as has been critically pointed out before, consideration of general audiences, which will inevitably involve people with different ability levels, is still in need of improvement in visualization research [44, 45, 32].

In 2017, Lee et al. [8] introduced a Visualization Literacy Assessment Test (VLAT) to counter a lack of solidified tools for measuring literacy. The test consists of 53 validated multiple-choice questions comprising data visualization tasks like value retrieval. It was implemented on 12 data visualizations and specialized in measuring performance for simple visualization tasks. Such tasks, whose solutions can be classified as correct or incorrect, are beneficial for a reliable assessment. They provide a solid measure but hardly allow for a complex depiction of how visually literate respondents are and perceive themselves. Focusing on these types of tasks means anchoring the assessment in one dominant ability field of VDL. This leads to disregarding other dimensions that have been proven to influence the understanding of visualizations. Notably, VLAT was recently broken down into a briefer version, the Mini-VLAT [9], to counter limitations of implementing an assessment structure, which originally demands 20+ minutes to be completed. While the briefer scope of the Mini-VLAT is more welcoming toward general audiences, it is still limited to correctly reading and retrieving values. The Critical Thinking Assessment for Literacy in Visualizations (CALVI) by Ge et al. [10] shifted its assessment focus to integrating critical ability as a part of Visualization Literacy. CALVI consists of 45 finalized items based on visualizations with multiple deliberate misleaders covering nine different chart types. The test takes a meaningful step towards a more complex ability understanding, but remains limited to an approach relying on data visualization tasks.

Data visualization tasks commonly used in prior literacy assessments like VLAT [8] or CALVI [10], such as value retrieval or data point comparison, primarily require readers to interpret visualizations in order to provide assessment answers. As discussed in Section 2.1, where ability categories from the learning sciences [13, 14] were related to VDL, engaging with data visualizations involves more than just interpreting. It can be argued that an assessment for VDL should extend beyond *Interpretation* to include abilities such as *Explanation*, *Application*, *Perspective*, *Empathy*, and *Self-knowledge* (see Figure 2.1).

2.3. Self-Perception in Ability Assessments

Recently, PREVis, a scale for perceived readability in data visualizations, was introduced by Cabouat et al.[46] and exemplifies how self-perception may provide insights into literacy. Using Likert scales, it measures four facets of perceived readability: understanding, visualization layout, data value readability, and perception of data patterns. PREVis is built on a strong theoretical foundation, but its exclusive focus on readability is limiting, as the authors acknowledge [46]. MAVIL overlaps with some PREVis aspects, such as perceived understanding (see MAVIL Dim1 in Table 2.1) and identifying visual elements (Dim2 in Table 2.2), but provides additional insights. For instance, our survey, detailed in Section 4, found that 22% of respondents struggled to comprehend "millimeters", and 28% found "millions of tons" unclear (Q2a).

Self-assessment assumes that human traits and skills are quantifiable and measurable [47] and can complement task-based tests to support more nuanced approaches to measuring abilities [48, 49]. In our study, self-assessment captures how visualization readers perceive facets of their VDL. We hypothesize that these self-perceptions can reveal nuanced insights into the ability and may correlate with objective performance on data reading tasks. Similar approaches have been widely applied in psychology, education and health research: self-concepts of talent have been compared to performance [50], self-perceptions have been linked to academic achievement [51, 52], perceived coping abilities have been shown to predict smoking cessetation outcomes [53]. Further, self-perception is recognized as a survey approach in political science, i.e. to gain insights into participants' stances and beliefs [54, 55].

While self-perceived ability ratings are not in the same way objective or precise like task-based performance measures, they can still be predictive [56]. Rather than replacing established methods like VLAT [8], we aim to integrate self-perceived ability ratings to enhance VDL assessments. Beyond measuring what people *can do*, self-assessment sheds light on how respondents *perceive* their interactions with visualizations. It remains underexplored how visualization readers' personal beliefs, trust and emotional responses – which have been shown to shape information retrieval [57, 58] – may affect visualization understanding.

2.4. Climate Data Visualizations as Assessment Objects

In developing a VDL assessment tool, we had to decide on the type of visual representation to present alongside the assessment questions. We chose to assess the comprehension of climate data visualizations, as they are meaningful for conveying scientific evidence about climate change and other current issues in research and the public discourse [15, 19, 20]. Though scientific evidence on climate topics should be broadly accessible, for example due to its interconnectedness with socio-political and cultural phenomena, it has been described as particularly challenging to convey. Technically, data visualizations are seen as suitable for communicating complex information effectively and efficiently [59, 11]. However, in reality the authors of the Intergovernmental Panel on Climate Change self-reflectingly discussed the complexities and tension between depicting data as scientifically accurate and comprehensible to wider audiences. They were cognizant that their own visualizations are not all easily understandable, identifying areas for improvement [21, 22]. Such improvements are needed to strengthen information distribution and trust in sciences related to the climate discourse, in which data visualizations play a crucial role [29].

For the development of MAVIL, a simple line chart and a simple bar chart with climate data (see Figure 1) were selected, assuming that a number of the surveyed audience would have encountered these types of data visualizations, e.g., in everyday news sources before. At the same time, others who might feel unfamiliar with the type of displayed visualization are likely to still find the examples comprehensible due to their simplicity. The choice of climate data exemplifies how content knowledge can be incorporated into an ability assessment, and exemplifies how displaying a particular topic may provoke a range of emotional reactions by visualization readers (see Section 4.1.5).

In the **line chart** (Figure 1.1a), the x-axis represents years, while the y-axis shows sea level height, starting from 0 mm in 1993 and reaching approximately 100.8 mm in 2021. A highlighted annotation in the middle of the graph states, "Sea levels are rising on average 3.5 mm per year," drawing attention to the steady upward trend. The line is shaded with a light blue area labeled "Divergence," emphasizing the rate of increase. In the horizontal **bar chart** (Figure 1.1b, CO_2 emissions of various countries in 1990 and 2018 are shown, measured in million metric tons. The bars are color-coded: dark gray represents emissions in 1990, while light gray represents emissions in 2018. A red or green percentage change is indicated for each country, showing the increase or decrease in emissions between these years. Each country's flag is displayed alongside its name, covering China, the U.S., India, Russia, Japan, Germany, Iran, South Korea, Saudi Arabia, and Canada. A large CO_2 cloud icon is placed in the middle of the chart.

2.5. Merging Ability Dimensions of Visual Data Literacy in MAVIL

Visual Data Literacy is considered complex and multilayered, and it can be defined drawing on varied expertise (as outlined in Section 2.1). Based on different approaches to VDL, the ability was deconstructed into six ability dimensions: *Comprehending*, *Decoding*, *Aestheticizing*, *Critiquing*, *Reading*, and *Contextualizing*. Each dimension informed the creation of a question block for the Multidimensional Assessment of Visual Data Literacy (MAVIL), which consists of self-assessment questions, text field prompts, and data reading tasks. The subordinate questions for each block are included below. The numerical order of dimensions reflects the order in which the question blocks were implemented in a survey format, and they should not be read in a hierarchical sense.

2.5.1. Dimension 1: Comprehending

First, self-perception of the overall graph comprehension, including abstract thinking, is surveyed. The perceived understandability of a data visualization by its reader can serve as an indicator of their potential comprehension. Consequently, the overall impression of the provided data visualization example is assessed (D1a, see Table 2.1). The complexity of visualized information – encompassing factors such as quantity, variety, and openness [22] – can significantly affect comprehension. Therefore, readers are asked to self-assess their ease of reading the data visualization (D1b). Additionally, in line with the definition of abstract thinking as markers of advanced visual literacy [33], participants are prompted to consider whether they could connect the shown information to their personal knowledge (D1c), and if they could summarize the visualization's content for others (D1d).

All questions in this initial block are self-assessment-based, with (partial) agreement to these statements theoretically indicating a high level of VD and suggesting a high likeliness of correctly understanding the examined data visualization. More detailed descriptions of the survey task design is provided in Section 3.1.1.

ID	Question in MAVIL	Theoretical foundation
D1a	The figure is understandable to me.	Introductory question to ask for subjective understandability rating.
D1b	I can easily read information from the visualization.	The complexity of visualized information may influence its understandability [22].
D1c	I can connect the shown information to my previous knowledge.	Abstract thinking may indicate a high level of visual literacy [33].
D1d	I could summarize the content of the visualization for someone.	

Table 2.1.: Overview of MAVIL questions on **Comprehending**, with question IDs, phrasing, and theoretical foundations.

2.5.2. Dimension 2: Decoding

The second set of questions focuses on graph-specific aspects, thereby examining bottom-up elements of VD from the perspective of the visualization reader. Dimension 2 focuses on self-perceived decoding of graph specifics, such as graph elements and familiarity with graphs. The comprehensibility of the displayed measurement units (D2a) and the clarity of color usage (D2b) are evaluated. This approach follows the idea that understanding visualizations depends on both structural elements like data units and visual aspects such as color [34]. Additionally, given that familiarity with graph types has been shown to aid in accurate data interpretation [11], this aspect is also explored (D2c, D2d).

All questions are presented in a self-assessment format (see Table 2.2). In this block, specific adaptations were made to the questions to suit the tested data visualization example and to enhance clarity for the intended general audience. For instance, a data

ID	Question in MAVIL	Theoretical foundation
D2a	I understand the units of measurement shown here, such as [insert example].	Visual literacy partly addresses understanding structural elements [34].
D2b	I can easily see if the colors used in the figure show information.	Visual literacy partly addresses understanding visual elements [34].
D2c	This is the first time I have read information from a [insert graph type].	Familiarity with a visualization type may increase the likeliness of correct graph understanding. [11].
D2d	I am familiar with [insert graph type] as a way to represent data.	

Table 2.2.: Overview of MAVIL questions on **Decoding**, with question IDs, phrasing, and theoretical foundations.

unit example referring the tested visualization is provided in D2a, and the shown graph type is specified in D2c.

2.5.3. Dimension 3: Aestheticizing

Dimension 3 addresses self-assessed aesthetic perception in relation to VD. Here, the *BeauVis* scale [35], which was developed in 2022 by He et al., was adapted (D3a-D3c). This scale consists of singular English-language adjectives that are synonymous and intended to capture various aspects of aesthetic perception. The descriptive terms of the *BeauVis* scale were translated into German, as this is the primary language of the surveyed population. This translation was performed in a language-sensitive manner to convey the original intent. The questions (D3a–D3c), as shown in Table 2.3, do not replicate the exact phrasing of the validated scale but are instead English translations of the German version we developed. The development of a German version was partly informed by a qualitative pilot study, further detailed in Section 3.2. The questions on aesthetic perception are supplemented by a question on perceived trustworthiness (D3d), as attributing beauty to a visualization may positively influence its perception [12]. All questions in this block are in a self-assessment format.

ID	Question in MAVIL	Theoretical foundation
D3a	The figure appeals to me.	A combination of adjectives can query the aesthetic perception of a data visualization (*BeauVis* scale [35]).
D3b	The figure looks very nice.	
D3c	This figure is well designed.	
D3d	The figure looks trustworthy.	Beauty may causally affect trust in visualizations [12].

Table 2.3.: Overview of MAVIL questions on **Aestheticizing**, with question IDs, phrasing, and theoretical foundations.

2.5.4. Dimension 4: Critiquing

Dimension 4 centers on self-perceived critical ability and understanding of visualization construction. Respondents are asked to provide specific feedback on the data visualization, including suggesting improvements (D4a, D4b), as literacy encompasses the ability to question and propose changes to a given object [38]. This aligns with the concept of the ability as a set of skills related to not only reading, but also writing, and creating visualizations [39, 60].

ID	Question in MAVIL	Theoretical foundation
D4a	I have a suggestion for a change to the figure.	Visual Literacy not only covers the ability to read, but also to write visuals [39].
D4b	If I could, I would design something in the figure differently.	
D4c	I would like to remark something positive about the figure.	Visual literacy involves critical inquiry and seeking improvement suggestions [38].
D4d	I would like to remark something negative about the visualization.	

Table 2.4.: Overview of MAVIL questions on **Critiquing**, with question IDs, phrasing, and theoretical foundations.

In this block, respondents were also provided with two optional text fields to articulate perceived visualization strengths (D4c) and weaknesses (D4d). These fields were optional, acknowledging that respondents with weaker writing skills may struggle to provide detailed responses [61, 62]. The questions in this block are presented in Table 2.4.

2.5.5. Dimension 5: Reading

Dimension 5 complements MAVIL with task-based questions by presenting data reading exercises (D5a-D5c), as this type of assessment question is well established, such as in VLAT [8] and CALVI [10], which were discussed in Section 2.2. These tasks directly relate to the data content of the visualizations being tested. The data reading tasks were presented in a true-or-false format. In Table 2.5 data reading questions specific to the tested bar chart (see Figure 1.1b) are provided as an example.

2.5.6. Dimension 6: Contextualizing

Finally, two contextual ability facets closely related to VD are surveyed: self-perception of numeracy skills and topic interest. A higher level of numeracy has been shown to indicate a greater likelihood of accurately reading information from a data visualization [36, 37]. Although established numeracy tests are effective, they would have been too lengthy for the concise assessment scope required here. Therefore, instead of using existing numeracy tests, respondents self-assessed their numeracy level (D6a, D6b) [48]. Both questions on numeracy are presented in a self-assessment format and can be found in Table 2.6.

ID	Question in MAVIL	Accuracy
D5a	According to the figure, CO_2 emissions are only declining in Russia and Germany over the course of 28 years.	True
D5b	The figure shows CO_2 emissions in billions of metric tons.	False
D5c	According to the figure, CO_2 emissions are increasing in most countries from 1990 to 2018. The smallest increase is in Japan.	True

Table 2.5.: Overview of MAVIL questions on **Reading**, with question IDs, phrasing of true-or-false statements, and indication whether statement is true or false. The data reading questions are specific to the bar chart shown in Figure 1.1b
.

ID	Question in MAVIL	Theoretical foundation
D6a	I have good math skills.	Numeracy skill is likely to increase the ability to understanding a data visualization [37].
D6b	I find it difficult to deal with statistical information.	

Table 2.6.: Overview of the first part of MAVIL questions on **Contextualizing**, with question IDs, phrasing, and theoretical foundations.

In addition to addressing numeracy as a personal skill, Dimension 6 explores personal attitudes toward the displayed content, as the understanding of visual representations has been shown to depend on the personal knowledge and interests of the readers [40, 11]. Therefore, four questions are tailored to assess interest in climate-related topics, as the visualizations used during testing featured climate data. This block includes questions on belief in climate change (D6c, D6d). These questions were developed for the *Digitize!* survey project, which aims to gather representative opinions and attitudes from the Austrian population, as further described in Section 3.4. All questions in Dimension 6 are presented in self-assessment format (see Table 2.7).

ID	Question in MAVIL	Theoretical foundation
D6c	Climate change is a proven fact.	Belief systems may shape understanding visual information [63].
D6d	The climate is not affected by human activity.	
D6e	I am concerned about climate change.	Reading accuracy is more likely if readers have expertise on a represented topic [11].
D6f	I deal with climate and environmental issues out of my own interest.	

Table 2.7.: Overview of the second part of MAVIL questions on **Contextualizing**, with question IDs, phrasing, and theoretical foundations.

3. Methodology

The development of MAVIL (Multidimensional Assessment of Visual Data Literacy) involved a design phase followed by three test implementation phases. The design of the assessment questions, described in Section 3.1, was based on the deconstruction of VDL into ability dimensions, as outlined in Background Section 2.5, and the questions were then formatted as a digital survey. In the first test phase (see Section 3.2), survey pilots were conducted with eight participants, and semi-structured interviews provided qualitative feedback to identify potential problems. After refining the assessment questions, the second phase (see Section 3.3) involved an online survey completed by 438 respondents, representative of Austria's population aged 18 to 74 years and the male-female gender split within those age groups. The data collected was used for quantitative analysis of the assessment, such as evaluating variable correlations. Finally, in the third phase (see Section 3.4), two blocks of MAVIL questions were included in the *Digitize!* social science survey [64], where 2373 respondents answered MAVIL questions. This phase further informed the development of the question blocks.

3.1. MAVIL in Practice: Design of a Survey

The design of MAVIL was based on the theoretical framework established in Section 2.5, where various approaches and studies related to VDL were analyzed to define multiple dimensions of the ability. This resulted in six question blocks, each with subordinate assessment questions grounded in the corresponding theoretical foundations, as presented in Tables 2.1 through 2.7. An overview summarizing the MAVIL question blocks, which were then used for test implementations, can be found in Figure 3.1. As shown in the figure, MAVIL comprises six ability dimensions of VDL, covered in six question blocks, along with a block for the demographic details of the visualization readers. All question blocks were formatted into a survey using *SoSci Survey*, a free software for the design and implementation of academic surveys [65].

When creating the initial survey blueprint, we targeted general audiences with no expected expertise in climate data or data visualizations. It was crucial to design with this audience in mind [8], to ensure that the assessment would follow the aim of being accessible. For example, short item phrasing was essential, as the survey format encourages relatively quick responses [66]. In each of the three testing phases, an iterated version of the MAVIL question catalog was presented to respondents. Thus, the test implementations are not directly comparable but build upon each other. The configuration of questions and the number of tested climate data visualizations varied across the phases of assessment testing.

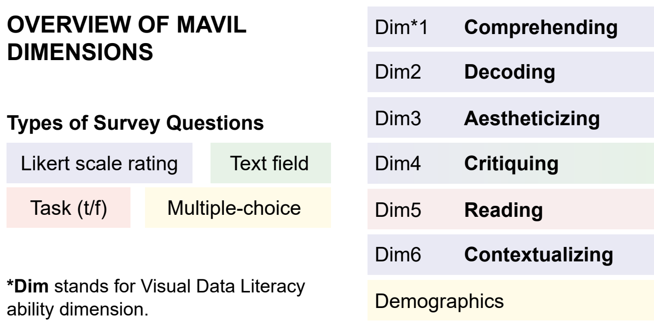


Figure 3.1.: Overview of MAVIL dimensions and their question types.

Please note that MAVIL was designed in German, as this is the primary language of the surveyed population, and testing was conducted with this language version. The assessment questions were translated into English for this thesis without testing the language adaptation. A full catalog of questions can be found in the appendix (see "MAVIL Catalog of Questions"), alongside the German-language original used during testing (see "MAVIL Fragenkatalog").

3.1.1. Survey Tasks

To cater to general audiences, there were no prerequisites for taking the survey. At the beginning of the assessment, respondents were made aware of the study's purpose of investigating the understanding of visually represented climate data. Respondents had to confirm their consent to participate in the study and that they were not underage. Instructions for completing the survey were kept concise but informative. In each test implementation, respondents were asked to engage with one or two data visualizations by paying attention to both content and overall impression. After initially viewing a data visualization, respondents could proceed to question blocks related to it. In each survey section directly referencing a data visualization, the corresponding chart was displayed beneath the assessment questions. The survey collected three types of data: ordinal data (e.g. self-assessment scale ratings), text data (e.g. text fields), and nominal data (e.g. true-or-false data reading tasks).

Likert Scale Ratings. Dimension1 on *Comprehending*, Dimension2 on *Decoding*, Dimension3 on *Aestheticizing*, Dimension6 on *Contextualizing*, and parts of Dimension4 on *Critiquing* consisted of self-perceived ability ratings questions (see Figure 3.2). These questions were presented as statements to be rated from the perspective of the visualization reader, and they were conveyed by an ordinal scale. Responses were given on a five-point scale, where 1 indicated "I disagree," 2 was "I somewhat disagree," 3 was "Undecided," 4 was "I somewhat agree," and 5 was "I agree." In the evaluation, ratings of 1 and 2 were grouped [67] to reflect disagreement with the statements, while 4 and 5 indicated agreement. A rating of 3 was treated as a neutral response. Based on the theoretical foundation of each question, either end of the Likert scale was interpreted as indicating high VDL or the likelihood of reading accuracy. The assumed interpretive meaning of agreeing with each self-assessment statement is summarized below for each question. For the theoretical foundations and full question phrasing, see Tables 2.1 through 2.7, as outlined in Section 2.5.

Text Fields. In Dimension4 on *Critiquing*, next to scale ratings, text fields were implemented. They asked for positive and negative points of visualization critique. The text fields were made optional to account for respondents who might have difficulty with writing, as this could result in briefer or less detailed answers in the survey [61, 62]. The fields had a restricted range of up to 2000 characters and provided text inputs were evaluated qualitatively. An optional text field for feedback was included at the end of all question blocks, allowing respondents to share any additional comments or suggestions about the survey. Providing a feedback option is considered good practice in survey design [68], as it offers valuable qualitative insights into the respondents' experiences and perceptions, which can highlight issues or ambiguities that may not be evident from quantitative responses alone [69]. This feedback field was optional and limited to 2000 characters.

Task-Based Questions. In Dimension 5 on *Reading* task-based data reading questions were presented in true-or-false format, as shown in Table 2.5.

Selection-Based Questions. For the demographic information section, selection-based questions were implemented. Demographic questions gathered age data using standard age ranges (in years) beginning with 18–24, 25–34, 35–44, 45–54, 55–64, 65–74, and 75 and older. Age ranges are a common practice in market research, and they were suggested by the market research organization which conducted an online survey with MAVIL questions (see Section 3.3). Age ranges allow for efficient grouping and analysis of respondents across standardized life stages. To ensure inclusivity, a variety of options were provided for gender identification, including "female", "male", "diverse", "inter", "open", "no gender entry", and an option to decline stating gender. Offering diverse options supports inclusivity by accommodating a range of gender identities, thereby making respondents feel recognized and respected [68]. The education level options were tailored to the Austrian education system, covering "no compulsory school certificate", "compulsory school certificate", "Matura", "apprenticeship/vocational training", "university degrees", and an option for degrees not represented in the listed choices.

3.1.2. Selection and Embedding of Climate Data Visualization Examples

Certain MAVIL ability assessment questions are effective in a survey only when directly linked to a visual representation. Consequently, data visualizations were selected for survey testing based on several criteria: they had to display climate-related information to align with the topic setting (for details on the choice of climate data visualizations, see Section 2.4). The displayed data was rather low in complexity, as to not overstrain respondents' willingness to engage with the visualizations. Since there were at times two data visualizations juxtaposed during testing, their chart types had to differ. The presentation format could not be interactive to keep the complexity low and clarify the respondents' engagement. Furthermore, all implemented data visualizations had Creative Commons licensing, allowing further public usage (CC BY-ND), and were credited accordingly in the survey display. In addition to these selection criteria, the visualizations had to be found suitable for presentation as assessment objects to general audiences. This was judged during short feedback exchanges with researchers from the Research Group of Visualization and Data Analysis at the University of Vienna.

In the first test implementation phase (the survey pilots) and also in the second phase (an online survey), each respondent encountered two data visualizations. Using two examples helped avoid bias from the unique characteristics of a single visualization. In an automated, balanced approach, random selection determined which visualization respondents saw first, preventing potential carryover effects [70]. This randomization was implemented in the *SoSci* software used for the survey administration, with functionality confirmed when survey data was downloaded and analyzed. In test phases where two visualizations were presented, some MAVIL question blocks appeared twice in the survey. This includes Dimension 1, Dim2, Dim3, and Dim4, as they directly relate to the tested visual representation. Data reading tasks in Dim5 were adapted to the content of each visualization example, which is why this block also appeared twice. Questions related to specific skills (Dim6), personal stance (Dim6), and demographics appeared only once in the survey.

3.2. Testing I: MAVIL Survey Pilots

Qualitative survey testing was carried out with eight respondents who filled out the assessment in the presence of a researcher. In accompanying semi-structured interviews, the respondents described their interpretation of the two data visualization examples and discussed their experience answering the assessment questions.

In the survey pilots, four climate data visualizations were used. For each pilot phase, with four participants each, two visual representations were tested. In the first phase of survey pilots, a line chart displaying temperature anomalies over the course of a year and a bar chart visualizing changes in pollutant emissions in Austria from 1995 to 2019 were shown. In the second phase, two additional data visualizations were tested to mitigate potential visualization biases: a line chart on rising sea level data Figure 1.1a and a bar chart on selected countries' CO_2 emissions over time Figure 1.1b. Since the tested visualizations were created for a German-speaking audience, the content of the latter two has been summarized in English (in Section 2.4) to improve understandability for this thesis, as their perception is discussed later on in more detail.

In this initial MAVIL testing phase, all question blocks (Dim1 to Dim6) were included, and demographic information was also collected. Dimension 3 on *Aeshteticizing* was introduced after the first phase of pilots had taken place, but was considered a substantial addition to include for the second phase of pilots. Note that the phrasing of MAVIL

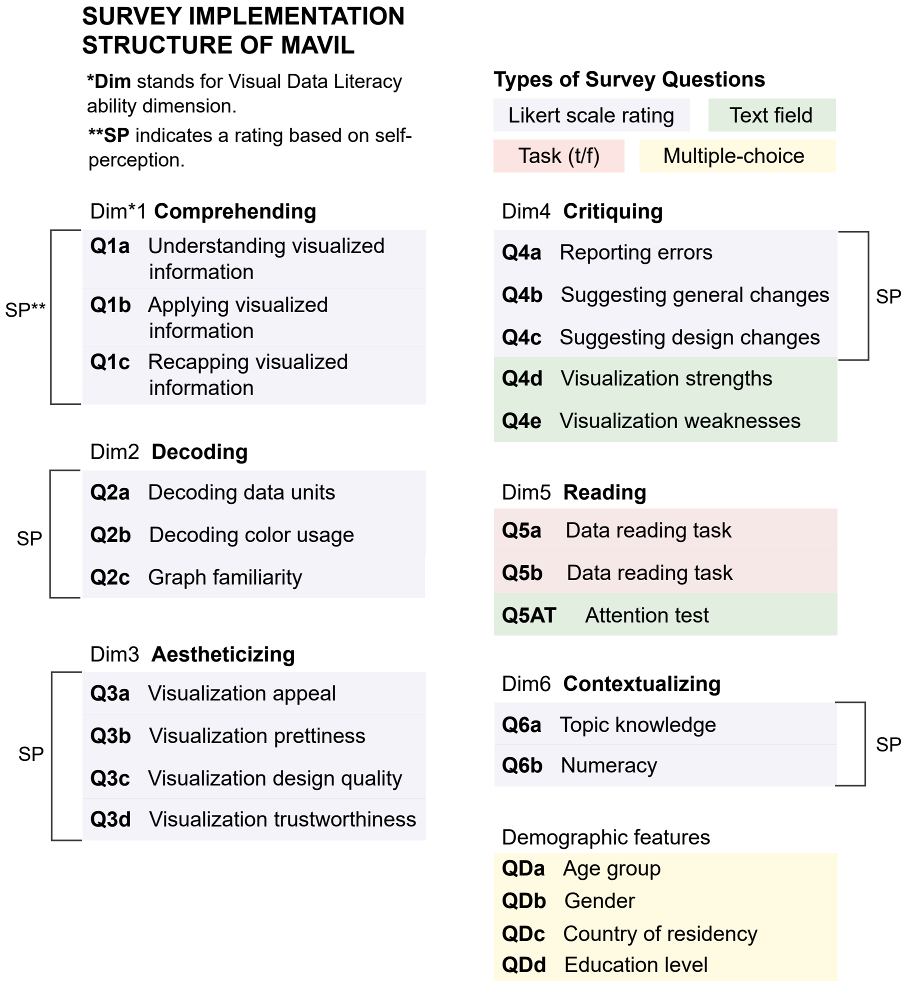


Figure 3.2.: Overview on MAVIL questions used in the pilots and the online survey: covering multiple VDL dimensions and surveyed demographics.

questions, as shown in Tables Table 2.1 through Table 2.7 in this thesis, differs from the version used in this testing phase in terms of the number of items per question block and specific wording. Since the questions were under development during the pilot phase, their exact phrasing is not displayed here but is summarized in a structural MAVIL overview in Figure 3.2. The full version of refined questions based on the survey pilot results and used for the second test implementation is included in the appendix (see "MAVIL Catalog of Questions").

3.2.1. Purpose and Procedure of Survey Pilots

The survey pilots supported sharpening and validating MAVIL questions. By analyzing the respondents' assessment experience, including question understanding, the feasibility of the questions could be qualitatively evaluated and then iterated upon. The insights were particularly productive for fine-tuning question phrasing. Participants were purposively recruited through professional and personal networks to ensure diverse educational and professional backgrounds. Of the eight respondents, six had no expertise in climate data visualizations, one was familiar with data visualizations in their field of work, and one was a visualization researcher. An overview of all eight participants is shown in Table 3.1.

The survey pilots were conducted in two phases of four pilots each. The first four pilot sessions (P1-P4) took place between December 10 and December 26, 2022. The second phase (P5-P8), after implementing initial revisions, was conducted from February 14 to February 27, 2023. All sessions were held in person and lasted between 28 and 47 minutes with a median duration of 35 minutes. Interviews were audio-recorded with participants' consent and then transcribed.

ID	Age range	Gender	Represented highest education level	Expertise related to data visualizations
P1	25 to 34	male	University degree	Work-related familiarity
P2	25 to 34	male	Univsersity degree	Visualization researcher
P3	18 to 24	female	Matura	none
P4	45 to 54	female	Degree not listed	none
P5	18 to 24	diverse	Apprenticeship	none
P6	45 to 54	male	Master craftsmanship	none
P7	25 to 34	female	University degree	none
P8	35 to 44	female	Apprenticeship	none

Table 3.1.: An overview of characteristics for the eight survey pilot participants.

3.2.2. Analysis of Interview Data from Survey Pilots and Learning Excerpts

The interview data was analyzed thematically by accumulating interview passages into themes and deriving a subsequent interpretation for assessment development. The analytical process took place in reference to Grounded Theory [71]. This qualitative method combines interview data into bundled theoretical interpretations justified by the subject matter. Interview data was coded and then assigned to categories [72] to gain an overview of respondents' assessment experience. The resulting themes from the interview data analysis are depicted with more detail on the eight survey pilots in Table 3.2. There, the applicability of subthemes to each survey pilot are broken down. The following two themes are exemplary as categories used for further survey development, which are based on the coded interview data:

Issues of Question Comprehensibility. In the first phase of survey pilots, all respondents pointed out unclear question phrasing, just as they encountered questions they had to reread to answer. This led to specific rephrasing during the first iterative revision. In the second phase of survey pilots, two respondents mentioned unclear phrasing, and stagnant rereading of questions only occurred in the case of one respondent. This observation, for example, led to a simplification of the data reading tasks in Dimension 5. In the survey pilots, the tasks were presented as multiple-choice statements reproducing the content of the data visualizations, with only one correct answer. This format was then changed to a true-or-false mode, with one statement on the data visualization appearing, whose accuracy had to be judged by the respondents (for an example see Table 2.5). It was deduced from comprehensibility issues identified during the survey pilots that this type of question might be more straightforward and less irritating for visualization readers, while still allowing us to gather data on their reading of the tested data visualizations.

Engagement With Text Fields. Dimension4 on *Critiquing* consisted of, among other questions, two optional text fields per visualization example. The survey pilots showed varying attitudes towards expression in text form. Three out of eight respondents principally skipped text fields; two articulated some thought; another two did not fill in the fields, even though they verbally expressed suitable inputs to the present researcher. The eighth respondent filled out all text fields conscientiously to, as was self-stated, perform well in the survey. So that no assumptions are made about the respondent's expressive ability or their response attitude, text inputs as indicators for critical ability were kept optional. Later in this thesis, the text inputs (Q4d/Q4e) are evaluated as additional insight into critical ability as a facet of VDL (see Section 4.1.5).

Theme	Subtheme (Participant Mentions)	Implications for Survey Development
Question Clarity	Question phrased unclearly (P1-P4, P7, P8)	Rephrase and simplify wording
	Question reread multiple times before answering (P1-P5)	(see above)
	Confusion about terms related to aesthetic perception (-/-, P6-P8)	Adjust German translation of *BeauVis* scale
Use of Optional Text Fields	Text fields generally skipped (P1, P3, P7)	Text fields may not engage all respondents, but provide valuable insights for some
	Brief thoughts articulated (P4, P8)	(see above)
	Text fields filled to "perform well" (P5)	(see above)
	No text field input, but relevant comments made in interview (P2, P6)	Enhance prompts for critical feedback by separating text fields from rating scales

Table 3.2.: The table shows themes and subthemes from the survey pilots with eight participants (P1-P8), alongside derived implications for survey development.

3.3. Testing II: Full MAVIL Application in a Survey

After the qualitative survey pilots, the assessment questions and survey design were refined to create a version of MAVIL validated through the pilots, and then presented to a representative survey panel. The panelists were representative of Austria's age group distribution from 18 to 74 years and the male-female gender split within those age groups. This survey was implemented with the aid of a market research organization. Valid survey cases had to pass certain quality criteria, as described below. The resulting data set, which will be evaluated subsequently (see Section 4), held 438 cases. The survey data set is available on Zenodo (here: `https://zenodo.org/records/12626652`), alongside a reference sheet for the data set and the questionnaire, all provided in English .

In the representative survey, two visualizations – previously tested in the qualitative pilots – were presented: a line chart on sea level data (see Figure 1.1a) and a bar chart on selected countries' CO_2 emissions over time (see Figure 1.1b) used for testing in the second phase of survey pilots. For the online survey test implementation, all MAVIL question blocks (Dim1 to Dim6) were included, and demographic information was also collected. Within Dimension 5 on *Reading*, two true-or-false questions were displayed for each visualization tested.

3.3.1. Data Collection in Cooperation With a Market Research Organization

So that MAVIL could undergo an implementation with representative scope, respondents were recruited in cooperation with a market research organization. The services of three different organization were compared and offers were obtained to carry out the assessment project. One of those organizations was chosen to guide respondents to the survey in a controlled manner and pre-selected them based on crossing the agreed-upon quotas of gender and age groups. The respondents originated from the organization's database or other websites promoting market research participation. Respondents received incentives, whose form depended on the panelist's affiliation with either the market research organization or other platforms. This cooperation was funded by the Vienna Science and Technology Fund (WWTF). The market research organization held demographic self-reported data from the panel, the accuracy of which was verified by collecting demographic data in the MAVIL survey.

During the cooperation with the market research organization, 457 respondents completed the assessment over 14 days via an online survey. The number of completed surveys was achieved in subsets of 50 completions. The data quality of the completed surveys was evaluated after each subset, followed by the criteria for valid survey cases as illustrated in the next subsection.

3.3.2. Quality Criteria for Valid Survey Cases

The completed survey cases were downloaded as a data set from the platform *SoSci*, where the survey had been self-administrated from our side. All 457 cases from the market research cooperation period from March 3 until March 17, 2023, were selected. A completed questionnaire held responses for all obligatory questions (i.e., excluding the optional text fields) and both tested data visualizations. Respondents who exited the survey early were omitted. Another 19 cases were disregarded, as these respondents failed to answer the attention question (Q5AT; see Figure 3.2) correctly. Referring to one of the data visualizations, this control question asked for naming a specific value, which could be read from the visualization. Explicitly, it asked to name the country first listed in an overview of selected countries' development of CO_2 emissions (see Figure 1.1b). The correct answer, in this case, was China. Acceptable inputs were "China", "china", "Chin" and the like. The resulting data set, consisting of completed surveys during which the attention test was passed, shows 438 valid cases.

3.3.3. Demographics of the Survey

Demographic information on the respondents included their gender, age group, and education level. Respondents also had to confirm that their country of residence was Austria. Inclusive options for gender self-assignment were offered (for detail see Section 3.1.1). Of the 438 respondents, 223 were male (50.9%) and 214 (48.9%) female. One respondent reported their gender as inter. Age groups were recorded in categories of 10-year ranges and represent the distribution of the surveyed population. The minimum registered age was 18 since the legal age was a consent criterion, and the maximum was 74 years, as this was the highest age of online panelists available to the market research organization.

The representation of education levels, though meaningful, was not sought after because, among other things, the standardized procedure of the market research organization did not provide for the representation of this demographic feature. The resulting educational backgrounds of respondents turned out to over-represent those with tertiary degrees, such as academic degrees, and under-represent those with primary degrees as the highest educational level. he educational background of the 438 respondents was as follows: 1.4% had no formal school degree, 5.7% completed compulsory education, 65.3% completed secondary education, 26.9% held a tertiary degree, and 0.7% indicated their educational degree was not listed. To accurately reflect the Austria's population, the distribution would require 17.3% with compulsory education, 63% with secondary education, and 19.2% with a tertiary degree. This distribution would then align with data from Austria's Federal Statistical Office in 2021 [73].

3.4. Testing III: MAVIL Module in the *Digitize!* Survey

As part of an academic collaboration, selected MAVIL components were tested within the *Digitize!* survey. This interdisciplinary project features different questionnaire modules, each lasting approximately three minutes, which are optional for the panel to complete. In the fifth wave of the *Digitize!* survey, MAVIL questions applied to the previously surveyed bar chart (Figure 1.1b) were included and answered by 2373 respondents during the survey's fifth conduction wave.

3.4.1. Survey Collaboration within the *Digitize!* Project

Digitize! Computational Social Sciences in the Digital and Social Transformation is a digitization project by multiple universities in Austria (University of Vienna, University of Linz, University of Graz and University of Salzburg). The project emphasizes ethical data collection and analysis within the social sciences [74], among other things by fostering the collaboration between social sciences, data science, legal studies and research ethics experts [64]. For example, computational text analysis in German was employed to explore the intersection of political party language and newspaper reporting [74]. A recurring online panel survey is a key component of the *Digitize!* project. The panel, representative of Austria, was recruited offline [74]. Between February 2022 and June 2023, four survey waves were conducted. The panel composition varies across waves, consisting of both returning participants and new respondents. Factors influencing the panel composition, such as incentive types, questionnaire entry points, and the length of the invitation letter, are systematically varied and documented [74].

An initial version of the MAVIL question module for the *Digitize!* survey was submitted in May 2023. This module was refined iteratively in collaboration with the *Digitize!* team until September 2023 and then collected survey data was provided by the team in December 2023. During this phase, MAVIL questions were administered by the *Digitize!* team rather than our side. As MAVIL was integrated into a broader survey, obtaining study consent and collecting demographic or personal information were the responsibility of the *Digitize!* team. Access to the wave five panel survey data was granted to the author of this thesis and their supervisors, allowing data analysis and review. However, sharing this data as part of the thesis is not permitted. Documentation and datasets for all survey waves are available for download on the open-access platform AUSSDA (The Austrian Social Science Data Archive)[1].

3.4.2. Creating a MAVIL Module Version for the *Digitize!* Survey

The exchange with the *Digitize!* team led to improvements in question phrasing, informed by the team's expertise and experience in survey conduction. To fit within the limited time frame of about three minutes for a *Digitize!* survey module, it was decided to present one data visualization to the panel, accompanied by three data reading tasks and an additional 12 MAVIL questions. These questions included items previously tested in an online survey (see Section 3.3), covering aspects from *Comprehending*, *Decoding*, and *Aestheticizing*. Through the inclusion of data reading tasks, *Reading* was also addressed. Given the survey project's focus on societal and social issues, the MAVIL dimension on *Contextualizing* was further developed. This involved creating a set of six questions on beliefs about and perceptions of climate change to query the personal interest of the survey audience more thoroughly than before. As discussed in detail in Section 4, a

[1]The archive AUSSDA can be accessed at: `https://data.aussda.at/`

notable number of respondents in the prior online survey had used text fields to express their opinions on climate change, highlighting the necessity of creating opinion-based items for the survey. The only ability dimension excluded from this survey module was *Critiquing*, as incorporating it would have exceeded the agreed upon time frame.

For all MAVIL questions in self-assessment format, alternative answer options were included in this testing phase. Since the core *Digitize!* questionnaire allowed respondents to select "I don't know" for self-assessment questions, this option was added to self-perceived ability ratings in the MAVIL question module.

The MAVIL module for the *Digitize!* survey consisted of four question blocks, summarized in Figure 3.3. **Question Block 1** queried beliefs about and perceptions of climate change, with six questions further developing the personal interest aspect of Dimension 6 on *Contextualizing*. **Question Block 2** was based on selected questions from prior survey testing, querying respondents' self-perception of ability dimensions. This block featured two questions each on *Comprehending*, *Decoding*, and *Aestheticizing*. All questions in Blocks 1 and 2 were rated on a five-point Likert scale ranging from 1 ("Do not agree at all") to 5 ("Totally agree"), with an option to select "I don't know.". **Question Block 3** contained three data reading tasks. These tasks were answered in true-or-false format, with statements about the visualization accompanied by the options "The figure provides exactly this information" or "The figure does not provide this information." **Question Block 4** asked respondents to rate their certainty about their answers to the data reading tasks on a scale from 0 ("Not certain at all") to 10 ("Totally certain"). The module's Question Blocks 2 and 3 were applied to a data visualization on climate data: the bar chart showing selected countries' CO_2 emissions over time (see Figure 1.1b), which had been tested in the second phase of the survey pilots and was previously used as one of two visualization examples in the online survey.

3.4.3. Survey Data Description for the MAVIL Module

The fifth survey wave of *Digitize!* took place from March 22, 2023, to December 3, 2023. Up to $2,373$ participants answered the MAVIL questions in an elective module to the core questionnaire. The module was conducted under the title *Klima* (German for "climate"). Around half of the respondents identified as male (49.6%) and the other half as female (49.8%), with 0.6% choosing not to disclose their gender. The group of respondents included individuals from all Austrian federal states and all age groups, ranging from 16 years to the "80+" category. Data weighting was applied to ensure representativeness across socio-demographic factors for gender, age, education, region, employment, marital status, migration background, religious affiliation and political factors [75]. This weighting allowed for a representative depiction of Austria's population.

MODULE IMPLEMENTATION STRUCTURE OF MAVIL
as part of *Digitize!* survey (wave 5)

Types of Survey Questions in Question Blocks
Likert scale rating | Task (t/f) | *Survey instruction*

Please indicate how much you personally agree with the following statements.
[randomize items]

QB*1	**Contextualizing**
QB1a	Climate change is a proven fact.
QB1b	The climate is <u>not</u> affected by human acitivity.
QB1c	I am concerned about climate change.
QB1d	I deal with climate and environmental issues out of my own interest.
QB1e	Many claims concerning climate change are exaggerated.
QB1f	There are more important things in life than climate change.

QB1 is based on MAVIL Dim6.

Please indicate how much you personally agree with the following statements.
[show visualization, randomize items]

QB2	**Comprehending**
QB2a	The figure is understandable to me.
QB2b	I have difficulty reading information from the visualization.
QB2	**Decoding**
QB2c	I am familiar with bar graphs as a way to represent data.
QB2d	This is the first time I have read information from a bar graph.
QB2	**Aestheticizing**
QB2e	The figure looks very nice.
QB2f	The figure does <u>not</u> look trustworthy.

QB2 is based on MAVIL Dim1, Dim2 and Dim3.

Please decide whether or not the illustration conveys exactly the information from the following statements.
[show visualization, randomize items]

QB3	**Reading**
QB3a	According to the figure, CO_2 emissions are only declining in Russia and Germany over the course of 28 years.
QB3b	The figure shows CO_2 emissions in billions of metric tons.
QB3c	According to the figure, CO_2 emissions are increasing in most countries from 1990 to 2018. The smallest increase is in Japan.

QB3 is based on MAVIL Dim5.

*QB stands for Question Block, used to differentiate these questions from other MAVIL questions identified as "Q1a" and so on.

Figure 3.3.: Overview of MAVIL questions in the module for the *Digitize!* survey: covering *Contextualizing*, *Reading*, and aspects of *Comprehending*, *Decoding*, and *Aestheticizing*. References to "MAVIL Dimensions" correspond to the dimensions covered in the full MAVIL version during the online survey implementation, as illustrated in Figure 3.2.

4. Findings

The survey data from two MAVIL (Multidimensional Assessment of Visual Data Literacy) test implementations are evaluated here. This includes an online survey conducted in collaboration with a market research organization with 438 respondents, representative of Austria's population aged 18 to 74 years and the male-female gender split within those groups (see Section 3.3). The data analysis from this implementation offers first indications of the feasibility of the assessment, and provides an overview of the population's ability composition across all VDL dimensions in MAVIL: *Comprehending*, *Decoding*, *Aestheticizing*, *Critiquing*, *Reading*, and *Contextualizing*. Additionally, it evaluates the perception of two surveyed data visualizations – a line chart (Figure 1.1a) and a bar chart (Figure 1.1b) displaying climate data.

The analysis primarily focuses on this first data set, but findings are further informed by a subsequent implementation of a MAVIL module in the *Digitize!* social science survey project with up to 2373 respondents (see Section 3.4). This phase featured a reduced selection of MAVIL questions tailored to the module format, emphasizing *Comprehending*, *Decoding*, *Aestheticizing*, and *Reading*, alongside a more detailed development of the *Contextualizing* dimension. The *Digitize!* survey responses, based on respondents' perception of the previously tested bar chart (Figure 1.1b), are also analyzed.

4.1. Insights from the Full MAVIL Survey

MAVIL was presented to an online survey audience by a market researcher in March 2023, where respondents answered 41 MAVIL items (for an overview see Figure 3.2 and for the full catalog of questions see "MAVIL Catalog of Questions" in the appendix) across two climate data visualizations, a line chart (Figure 1.1a) and a bar chart (Figure 1.1b). As described in Section 3.3.2, in accordance with quality criteria for valid survey cases, there were 438 respondents who took a full version of MAVIL in appliance to two data visualizations.

All statistical analysis was conducted in Python with the statistical tests provided within the *scipy.stats* module. The correlations (in Section 4.1.3) were calculated according to Pearson's correlation coefficient using the median whenever results had to be aggregated over dimensions or visualizations. For the binary results (in Section 4.1.4), both a t-test for the arithmetic mean and a moods median test were conducted and gave similar results. Both tests gave the same result with regard to rejecting the null-hypotheses.

4.1.1. Insights into Performance Measures

To give a general impression about completing the assessment: It took respondents an average of 2.75 minutes to answer the main body of MAVIL questions per data visualization. On average, framing elements of the survey, such as the introduction and the query of demographic info, took about a minute to be completed.

Based on the data reading tasks posed in MAVIL Dimension 5, there were clear deficits when evaluating the surveyed population's VDL from a task-based point of view. Overall, 48% of the respondents made mistakes when reading the simple line and bar chart. Effect sizes of differences between the two charts are relatively small, as seen when comparing the answer distribution of correct answers (see Figure 4.1). The results of the data reading tasks in Dimension 5 across the visualizations were summarized into a Data Reading Score (DRS), which corresponds to the sum of all correct data reading answers per respondent. In Table 4.1 grouped correct answers are summarized, showing that, for example, about half (52%) of all participants had answered all four questions correctly. And only 5% of respondents answered only one data reading task total correctly.

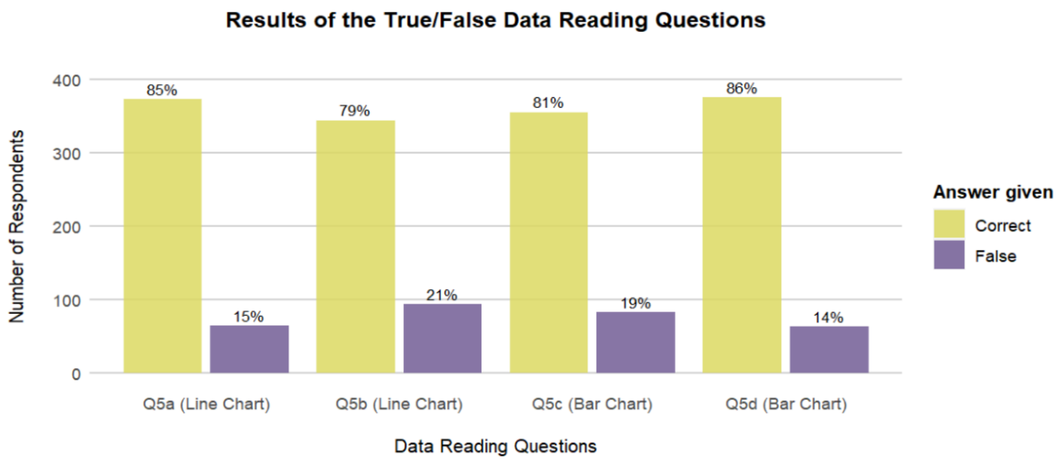


Figure 4.1.: Tasks for *Reading* were presented in true-or-false mode in this survey implementation. In the chart, the percentages of correct and false responses given are shown, the total number of answers being 438.

	Line Chart 2 Correct	Line Chart 1 Correct	Line Chart 0 Correct
Bar Chart 2 Correct	52%	18%	3%
Bar Chart 1 Correct	13%	5%	2%
Bar Chart 0 Correct	4%	3%	-

Table 4.1.: The table shows respondents (in %) with correct answers for the *Reading* questions, indicating that roughly half answered all questions correctly.

4.1.2. Respondents' Self-Perceived Abilities and Perception of the Surveyed Charts

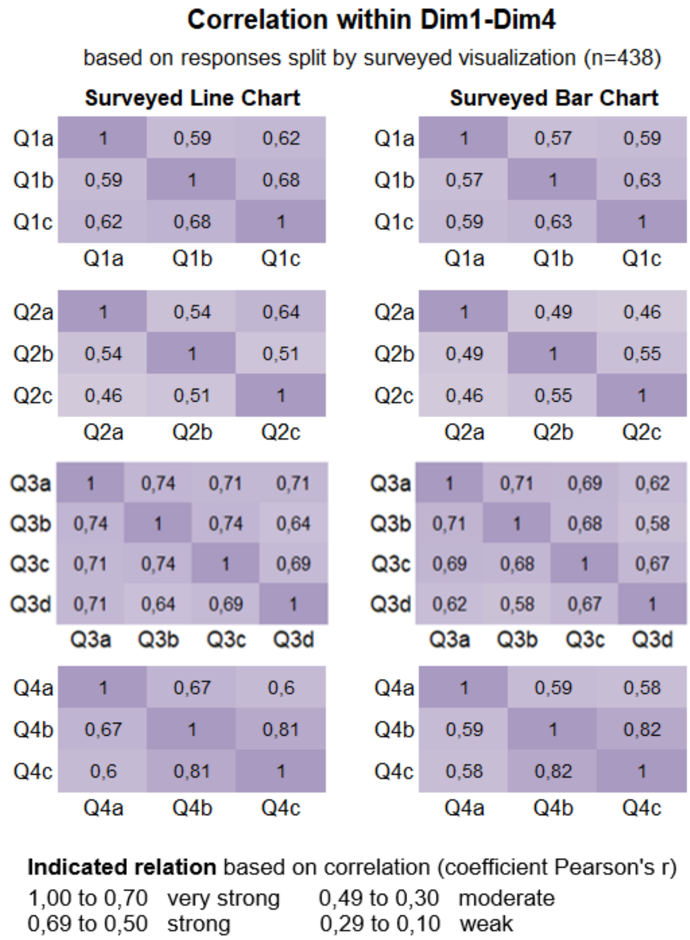
Although deficits are notable regarding *Reading* as an abiltiy dimension, the task-based measures do not uncover causes for visualization understanding deficits. Through a look at self-assessement questions, the ability composition of respondents mirrors certain deficits – which may be connected to the data reading performance – more distinctly. In contrast to the data reading performance, the self-perception of *Comprehending*, which covered a general perception of having understood the visualization, being able to apply and summarize its information, differed. A considerably smaller portion of the respondents who made mistakes in the data reading tasks indicated that they are **not** able to summarize the content of the shown data visualizations (Q1c; for an overview of all questions see Figure 3.2). In the case of both charts, only 5% of respondents stated that they cannot summarize the visualization. This suggests the possibility of misleading information transfer, as many respondents are unaware of reading deficits.

In the ability dimension on *Decoding*, the perceived ability to decode data units, color usage and rate graph familiarity were addressed. Regarding the displayed chart types (Q2c), 20% of respondents did not indicate familiarity with the line chart, and 19% did not indicate familiarity with the bar chart. The intersection of these statements results in 12% of respondents who felt unfamiliar or uncertain with the line and the bar chart as graph types. The survey also showed deficits in understanding simple measurement units (Q2a). 22% of respondents did not fully comprehend the usage of millimeters, and 28% of respondents found millions of tons somewhat incomprehensible.

There was a notable difference in the aesthetic perception of the surveyed climate data visualizations. This was queried in *Aestheticizing*, which covered visualization appeal, prettiness, design quality and trustworthiness. Around 10% of respondents rated the impression of the bar chart more positively than that of the line chart. In particular, 60% agreed to find the surveyed bar chart pretty (Q3b), and 72% found it well designed (Q3c). The line chart was perceived as slightly less aesthetic with 58% agreeing to find it pretty, and 68% finding it well-designed. This information tends to agree with the assessment of visualization trustworthiness (Q3d), as 75% of respondents found the bar chart trustworthy, with a negative five percent difference from those who found the line chart trustworthy.

4.1.3. Correlative Exploration of MAVIL Variables

The correlative exploration of ability dimensions and subordinate questions critically informs future work on the assessment structure by uncovering relations between ability dimensions and questions within those dimensions. The exploration highlights some variables, such as self-assessed numeracy, as significant indicators for VDL. The correlative exploration describes the relation of MAVIL dimensions to one another and the internal relations of single questions within an ability dimension. This is done by applying statistical methods and significance tests. The findings from exploring correlative relations in the survey data set do support the creation of self-assessment questions for different ability dimensions, as questions within Dimensions 1 through 4 show steady inner cohesion (see Figure 4.2) and simultaneously, these dimensions cover distinctive ability areas (see Figure 4.3). These four dimensions on *Comprehending*, *Decoding*, *Aeshteticizing*, and *Critiquing* are highlighted in the correlative exploration due to a similar number of self-assessment format questions in each dimension.



Correlation within Dim1-Dim4

based on responses split by surveyed visualization (n=438)

Surveyed Line Chart

	Q1a	Q1b	Q1c
Q1a	1	0,59	0,62
Q1b	0,59	1	0,68
Q1c	0,62	0,68	1

	Q2a	Q2b	Q2c
Q2a	1	0,54	0,64
Q2b	0,54	1	0,51
Q2c	0,46	0,51	1

	Q3a	Q3b	Q3c	Q3d
Q3a	1	0,74	0,71	0,71
Q3b	0,74	1	0,74	0,64
Q3c	0,71	0,74	1	0,69
Q3d	0,71	0,64	0,69	1

	Q4a	Q4b	Q4c
Q4a	1	0,67	0,6
Q4b	0,67	1	0,81
Q4c	0,6	0,81	1

Surveyed Bar Chart

	Q1a	Q1b	Q1c
Q1a	1	0,57	0,59
Q1b	0,57	1	0,63
Q1c	0,59	0,63	1

	Q2a	Q2b	Q2c
Q2a	1	0,49	0,46
Q2b	0,49	1	0,55
Q2c	0,46	0,55	1

	Q3a	Q3b	Q3c	Q3d
Q3a	1	0,71	0,69	0,62
Q3b	0,71	1	0,68	0,58
Q3c	0,69	0,68	1	0,67
Q3d	0,62	0,58	0,67	1

	Q4a	Q4b	Q4c
Q4a	1	0,59	0,58
Q4b	0,59	1	0,82
Q4c	0,58	0,82	1

Indicated relation based on correlation (coefficient Pearson's r)

1,00 to 0,70	very strong	0,49 to 0,30	moderate
0,69 to 0,50	strong	0,29 to 0,10	weak

Figure 4.2.: Correlation table of MAVIL questions (Dim1-Dim4) split by surveyed visualization example. All correlations of questions are statistically significant with p-values below $0,05$.

A consistent positive correlation of self-assessment questions within the dimensions on *Comprehending*, *Decoding*, *Aestheticizing*, and *Critiquing* indicates the plausibility of the construction of ability dimensions. As can be seen in the correlation tables in Figure 4.2, the relations between the questions within each dimension, and for both surveyed data visualizations, are fairly strong. For Dimension 1 on *Comprehending* (Q1a-Q1c) and also for Dimension 2 on *Decoding* (Q2ac-Q2c), all correlative coefficients are moderate $(0,3\text{-}0,49)$ to strong $(0,5-0,69)$. The correlative coefficients within Dimension 3 on *Aestheticizing* (Q3a-Q3c) and Dimension 4 on *Criticizing* (Q4a-Q4c) are strong or very strong $(0,7\text{-}1)$. Note that for Dimension 4 on *Critiquing* only self-assessment format questions (Q4a-Q4c) are included in the corelative analysis, and optional text fields (Q4d, Q4e) are evaluated separately (see Section 4.1.5).

Correlation of Dim1-Dim4 and DRS

(grouped responses of two surveyed visualizations, n=438)

	DRS	Dim1	Dim2	Dim3	Dim4
DRS	1,00	0,30	0,27	0,18	-0,17
Dim1	0,30	1,00	0,65	0,66	-0,27
Dim2	0,27	0,65	1,00	0,53	-0,18
Dim3	0,18	0,66	0,53	1,00	-0,24
Dim4	-0,17	-0,27	-0,18	-0,24	1,00

Indicated relation based on correlation coefficient Pearson's r

1,00 to 0,70	very strong	0,29 to 0,10	weak
0,69 to 0,50	strong	-0,10 to -0,29	weak, negative
0,49 to-0,30	moderate		

Figure 4.3.: Correlation table of Dim1 to Dim4 and the Data Reading Score (Dim5), grouped across visualization examples. All correlations are significant with p below $0,05$.

Following the correlative exploration of answers within dimensions for each surveyed data visualization, survey answers for Dimensions 1 through 4 were fully grouped across questions and visualizations. This is visualized in a correlation table in Figure 4.3, showing that that Dimension 1 on *Comprehending* strongly correlates with Dimension 2 on *Decoding* and Dimension 3 on *Aestheticizing*. Their correlation with the Data Reading Score (DRS) which corresponds to the sum of correctly solved *Reading* tasks per respondent, varies and remains weak $(0,1\text{-}0,29)$ to moderate. Dimension 4 on *Critiquing* disrupts this dynamic, as it is the only dimension correlating negatively with Dimension 1 through Dimension 3 and the DRS. This could be due to the simplicity of the data visualizations, as they might limit the scope of finding errors (Q4a) or coming up with design suggestions (Q4c). Respondents might react differently if hidden errors or intentional design issues were to be part of the assessed visualizations.

4.1.4. Significance of Age and Numeracy

In the survey implementation, selection-based demographic questions were posed to the respondents, asking for age, gender, education level and country of residence. Out of those features, age was deemed to significantly relate to VDL, in particular *Reading*. Out of the MAVIL questions in self-assessment format, numeracy as part of *Contextualizing* was a significant feature to describe the surveyed population. Self-perceived numeracy correlated significantly with the number of correctly solved data reading tasks.

Binary t-tests and Mood's median test on MAVIL variables, including demographic features, were carried out in relation to the DRS, which could range from 0 to 4 correct answers (see Table 4.1). The tests showed that the youngest age group of the surveyed population outperformed the older respondents regarding their *Reading*. The age group from 18 to 24 years (14% of respondents) had a DRS median of 4 (with $p=0,01$). The age groups from 25 to 74 years (86% of respondents) had a DRS median equal to 3 (with $p=0,01$). Additionally, the self-assessed numeracy level (Q6b) correlates significantly with the *Reading* performance. Those respondents who rated their numeracy level high (37% of respondents) tended to have a DRS median equal to 4 (with $p=0,009$). Those who rated their numeracy level neutral or low (63% of respondents) had a lower DRS median equal to 3 (with $p=0,009$). This suggests the informative value of having respondents self-assess their ability level.

There is no correlation between the highest level of education reported and the self-assessed level of numeracy (Pearson's $r=0,016$). As expected, our analysis did not observe a meaningful difference in VDL dimensions based on gender.

4.1.5. Report on Visualization Criticism in the Text Inputs

For each visualization, respondents had the option to provide written visualization criticism in two text fields (max. 2000 characters each), one for visualizations strengths (Q4d) and one for visualization weaknesses (Q4e). These questions were not mandatory, yet 37% submitted at least one meaningful response. All inputs for the line chart and the bar chart were thematically analyzed [71] and categorized into i. Visualization Strengths, ii. Visualization Weaknesses and iii. Climate Commentary.

Categories of Text Inputs and Examples

i. Visualization Strengths were queried in Question Q4d and defined in our analysis as a category of text inputs where a clear positive point of critique related to the tested visualization was made. For the Line Chart 99 inputs were given, such as finding the chart "inviting, easy to understand, the color blue well chosen", or that noting it was "kept quite simple not cluttered". Someone else highlighted that "the source is indicated, [which] must be highly credited". For the bar chart 113 according inputs were provided, describing it as "easy to understand and memorize", acknowledging that it "raises awareness", or finding "described data easily comparable".

ii. Visualization Weaknesses could be articulated for Q4e and in our analysis were defined as comments highlighting clear negative points of critiques related to the tested visualization. For the line chart, 87 inputs were provided, stating, for example, that "the graphic looks outdated, could be designed better". A few respondents were "missing axis labeling" or wondered what the uncertainty margin means. For the bar chart there were 78 inputs, including debate on the cloud icon, finding it "unnecessary" or "strange" (though this graph element had also been highlighted as a strength, e.g. someone said that the "cloud speaks to me"). Others criticized the comparison of countries, and, for instance, found the "relation to population missing" or wanted more information because "there are other countries worth mentioning that also emit high levels of co2".

iii. Climate Commentary appeared in Q4d and Q4e and emerged in our analysis as opinions about and reactions to climate change, rather than comments on tested visualization itself. In total, 35 of such input were given. To give an impression of this text input category, one respondent wrote that "the so-called man-made climate change is pure fraud", others called climate change "a lie" and one respondent said that the visualization would aim towards "tricking people into fear". Others articulated feeling affected such as that the visualization "makes [them] sad" or that looking at it they felt that "the world will be destroyed".

Statistical Analysis of Text Inputs

A chi-squared test was employed to evaluate the relation between the non-ordinal text inputs with demographic factors and respondents' Data Reading Scores (DRS). For this test, i. Visualization Strengths (212 inputs total) and ii. Visualization Weaknesses (165 inputs total) were considered equally by coding them as 1s and putting them in a 2x2 contingency table for each data visualization. Due to partially small sample sizes, significant findings from the chi-squared test were validated through an additional Monte Carlo simulation [76, 77] by randomly sampling $2,000$ replicates.

For both the line and bar chart, results indicated a significant difference in text input provision across education levels (Line Chart: X^2=12.93, df=1, p=0.0003, Bar Chart: X^2=7.88, df=1, p=0.005). The simulated p-values from the Monte Carlo simulation, based on 2000 replicates, validate the significance of this finding (Line Chart: p=0.0005, Bar Chart: p=0.005). This suggests, respondents with secondary and tertiary educational degrees were the ones who tended to fill in the optional fields on visualization criticism. As educational disparities can impact survey responses [78], possibly, higher-educated respondents may be more accustomed to complex survey items and have a structural advantage in providing written survey answers.

Moreover, a significant association was found between the provided text inputs and the DRS (Line Chart: X^2=10.40, df=1, p=0.001, Bar Chart: X^2=8.91, df=1, p=0.003). This association was validated through a Monte Carlo simulation with 2000 replicates (Line Chart: p=0.001, Bar Chart: p=0.002). These findings indicate, that the respondents with a higher DRS of 3 or 4 were the ones who were more likely to articulate visualization criticism. This underscores the position, that high critical ability is linked to a higher level of VDL [38].

In the chi-squared tests, no significant differences were found regarding the association of text input provision across neither age groups nor gender.

The inputs from Category III, which are statements of opinion regarding climate change, were not part of statistical testing. Yet, they are important to note, as they mirror the high salience of the climate change discourse and the connected emotional connotations triggered by the display of climate data. These inputs show opinions on climate change and often, the factualness of climate data was denied or the relevance of the survey was dismissed. This can be valuable when analyzing respondents' engagement with visual information displays, as their emotional state can influence how information is extracted [79]. The input in Category III motivated the design of a more elaborate assessment of belief in and personal interest in climate change as part of Dimension 6 on *Contextualizing*. This was carried out during the subsequent survey implementation of a MAVIL module, which is described in Section 3.4 and whose results are presented in the next Section 4.2.

4.2. Insights from the MAVIL Module in the *Digitize!* Survey

A selection of 15 MAVIL questions was presented as an optional survey module as part of the *Digitize!* social science project. In this implementation (described in Section 3.4), 2373 respondents respondents aged 16 to 80 and above, with an equal distribution of male and female gender identifications, answered the questions between September and December 2023. This implementation focused on whether attitudes toward climate change – addressed in *Contextualizing* – influence respondents' understanding of a climate data visualization. The module questions are summarized in Figure 3.3 and refer to the previously tested bar chart on CO_2 emissions over time (Figure 1.1b).

The survey data was weighted using values provided by the *Digitize!* team to adjust the sample distribution to Austria's total population, addressing measurement errors in the raw data. This ensures that groups such as young people or individuals with higher educational backgrounds, who are often overrepresented in survey samples, do not disproportionately influence the results. The following insights were developed in collaboration with the *Digitize!* team for a brief report [80] on our question module.

4.2.1. Contextualizing: Focus on Personal Attitude to Climate Change

The majority of respondents in Austria agree that climate change is partly man-made but report feeling less personally concerned. Attitudes toward climate change are summarized in Figure 4.4. Approximately three-quarters of respondents (69.9%) agreed that climate change is influenced by human activity (QB1b; see Section 3.4 for an overview of MAVIL module question), while 12.3% disagreed or strongly disagreed. Over half of respondents (57.3%) expressed interest in climate-related or environmental topics (QB1d), and 60.5% indicated concern about climate change (QB1c). A smaller proportion (15.3%) reported being somewhat or not at all concerned. Meanwhile, 26.7% believed that many statements

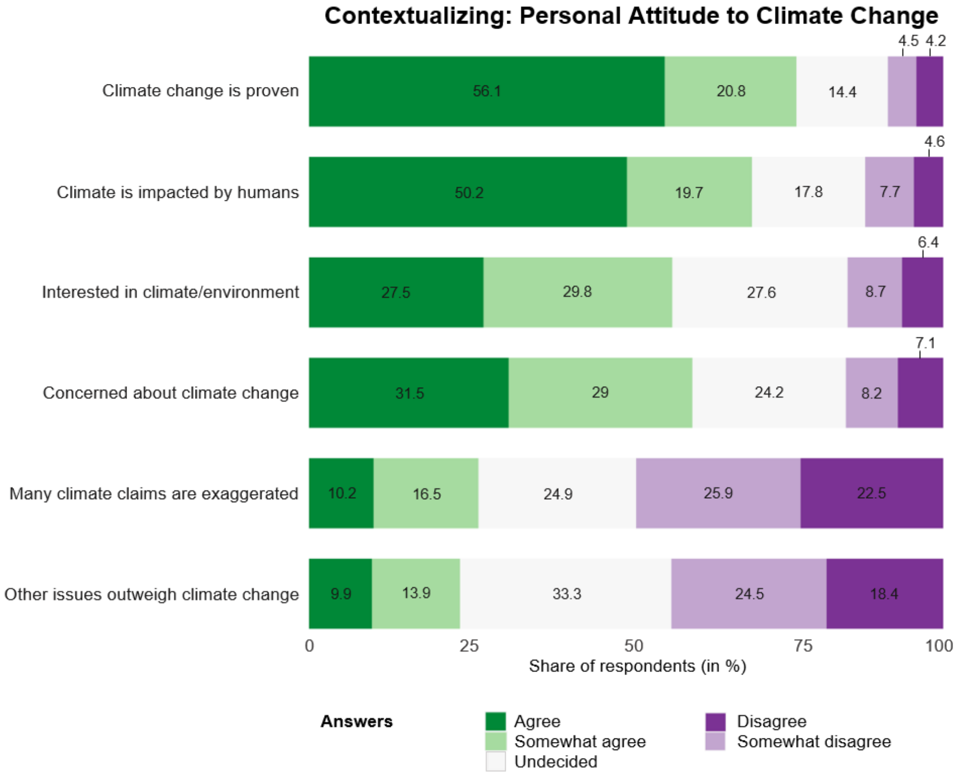


Figure 4.4.: The figure shows the response distribution for questions QB1 (n=2,373). The answers are weighted. To facilitate readability and due to the low number of "Don't know" responses, this answer option is not shown.

about climate change are exaggerated (QB1e), while 48.4% disagreed.

4.2.2. Reading, Comprehending, Decoding and Aestheticizing

For about a third of respondents, reading climate data from the surveyed bar chart proved challenging. Approximately 40% were unable to answer at least one data reading question (QB3a-QB3c) correctly. Respondents who agreed that climate change is influenced by human activity were more likely to correctly interpret the climate data. Regarding comprehension, most respondents (83.6%) found the bar chart understandable (QB2a), and only 5.4% reported difficulty reading information from it (QB2b). Similarly, most respondents (74.4%) stated they were familiar with bar charts (QB2c), while only 6.2% indicated that they were encountering this type of visualization for the first time (QB2d). Despite just under 10% abstaining, the chart was rated as trustworthy by a majority (67.2%, QB2e) and aesthetically pleasing by 51.9% (QB2f). Figure 4.5 summarizes respondents' self-assessments of selected VDL ability dimensions.

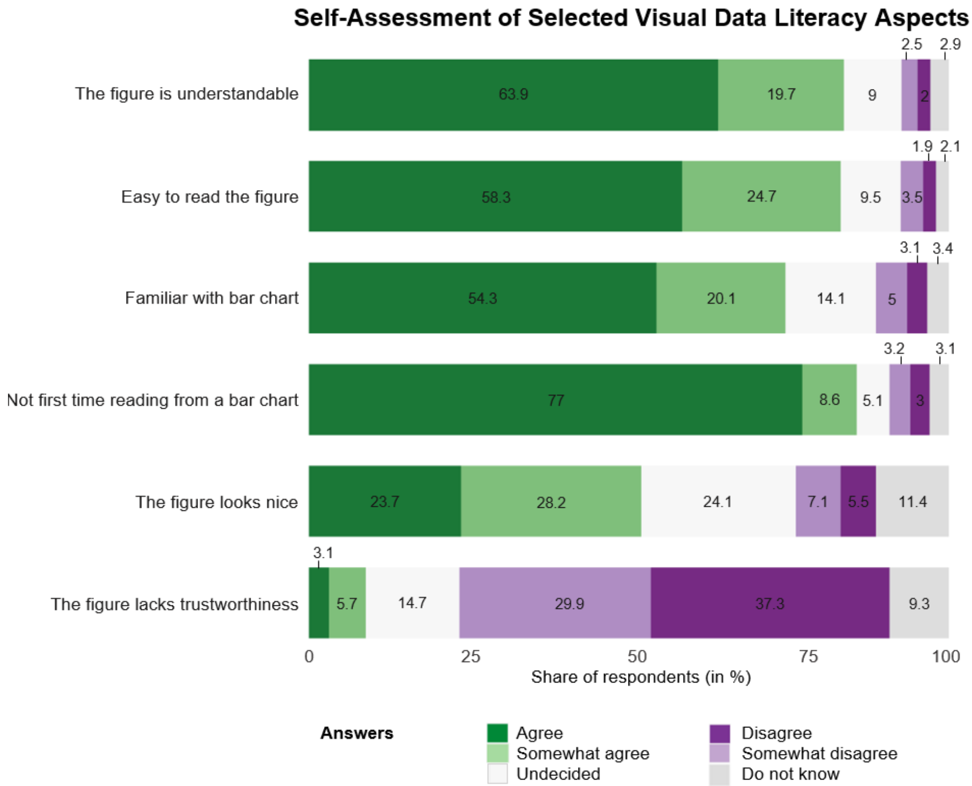


Figure 4.5.: The figure shows the response distribution for questions in QB2 (n=2, 360). The answers are weighted.

5. Discussion and Conclusion

With MAVIL, a broader lens for assessing Visual Data Literacy is proposed – one that acknowledges both performance and perception. While our goal is not to replace tools like VLAT [8] or CALVI [10], MAVIL challenges established assessments by introducing additional ability dimensions and integrating underutilized question types, such as self-perceived ability ratings and open-text critique. This multidimensional approach offers insights into how general audiences perceive and relate to data visualizations – insights that task-based methods alone cannot fully capture.

Here, we reflect on the conceptual and practical limitations of our case-specific implementations of MAVIL, discuss what our survey results reveal about different dimensions of literacy, and propose directions for expanding MAVIL's applicability. Although the focus during survey testing lay on two simple, static climate visualizations, the results offer a compelling entry point into better understanding how audiences engage with real-world data visuals and what assumptions researchers, designers, or anyone who creates or shares visualizations might need to rethink. The discussion reflects on key findings, highlights limitations in conceptualizing and conducting the assessment, and outlines possible future research endeavors. Finally, the contributions arising from the assessment's design and the insights gained from its evaluation provide a conclusion to this thesis.

5.1. Survey-Based Insights from MAVIL

Traditional assessments [8, 10] excel in task-based performance, such as value retrieval or pattern identification, but overlook other dimensions of how audiences perceive and interpret visualizations. MAVIL broadens this view by integrating six theory-driven dimensions of Visual Data Literacy (VDL), including self-perceived abilities. This allowed us to capture subtleties related to VDL and visualization understanding that established tools cannot provide.

During MAVIL testing, we gathered representative survey data (N1=438, N2=2373) and collected insights into two climate data visualizations (see Figure 1) across six VDL dimensions. We assessed not only performance through data reading (MAVIL Dimension 5), but also perceived comprehension (Dim1) and decoding ability (Dim2), aesthetic judgment (Dim3), critique behavior (Dim4) and context-related self-assessments like numeracy and topic interest (Dim6). While findings are specific to the visualizations used during testing, they offer valuable implications for information design and the understanding of critical data by general audiences. Even simple, static charts still pose challenges, which is exemplified by 8% to 12% of participants feeling unfamiliar with either bar or *both* line and bar charts. The correct understanding of data visualizations cannot be assumed and instead friction points in audience's sensemaking are worth noting and exploring further [81].

The full MAVIL survey further revealed that roughly a quarter of participants felt unable to decode data units like millimeters and millions of tons, highlighting that elements implicitly assumed to be self-explanatory by designers or curators [82] may require reconsideration as their correct interpretation by general audiences cannot be assumed. This challenges one-size-fits-all approaches, which have been criticized for being ineffective with diverse users who have different literacy levels and types of knowledge [83, 84]. Respondents' written feedback further illustrated this point. In Dimension 4 during the full survey, participants provided a total of 377 points of critique on visualizations strengths and weaknesses, with some comments highlighting distracting design choices or missing labels, offering grounded insight into what viewers notice – and what may hinder understanding. However, most participants did not feel confident suggesting design improvements in their self-ability ratings, indicating a gap between recognizing and articulating critique. It is likely that more flawed or intentionally misleading visuals (as for instance done by [85] or [10]) would elicit stronger responses, and future assessment implementations could explore this space.

Including aesthetic perception in our literacy assessment revealed that participants who found visualization visually appealing also tended to trust it, suggesting that, line with existent literature [86], design and perceived credibility are closely linked. Understanding this dynamic more can improve how we assess whether essential information is effectively retrieved from data visualizations [23, 24].

One key lesson from the test implementations is the importance of accounting for opinion expression in highly salient topics like climate change. During the full MAVIL survey, we observed that some respondents held strong views on climate change (see Section 4.1.5), which were not initially accounted for in our design. The Digitize! test implementation addressed this by expanding the *Contextualizing* dimension (see Figure 3.3) to better capture personal attitudes toward climate data. Merging these refinements with the original MAVIL design would enhance future assessments by ensuring that both factual comprehension and subjective perspectives are meaningfully integrated.

5.2. Limitations

There are representation limits for the surveyed population. Respondents below the age of 18 and above the age of 74 were not considered in the survey because of legal age as a consent criterion and due to availability reasons (see Section 3.3.3). In this test implementation, respondents shared the commonality of being incentivized by a market research organization, which may introduce a bias toward individuals with a frequent online survey participation and therefore higher digital skill. However, the approach we have chosen to select respondents is also standard in other representative surveys (e.g., [87, 88]). We recognize, that depending on the socio-cultural and geographical background of respondents, the assessment data would vary. However, we see these possible variations as a strength of our multidimensional assessment approach, as we would expect audience's perception to vary. MAVIL is open to different perspectives, as by allowing respondents to choose whether they want to express themselves in written text, and could potentially be a flexible framework for cross-cultural or cross-contextual applications.

MAVIL's integration of self-perceived ability ratings offers valuable insights into how people engage with visualizations, but it also introduces sources of noise. Self-assessments can be influenced by factors like over- or underestimation, social desirability, or ambiguity in question interpretation [52, 89]. While we have piloted MAVIL in several qualitative rounds and payed attention to straightforward question interpretation, such biases remain inherent to the method.

The visualizations used in the survey, a line chart and a bar chart on climate data, were chosen for their visual simplicity and assumed broad familiarity. This helped ensure accessibility to general audiences but may limit the scope of our insights. These visualizations are not representative of the type variety, complexity levels, and design possibilities of data visualizations. MAVIL should be tested with more diverse visualizations, such as scatterplots or intentionally misleading ones, to evaluate how it performs under varying complexity and design conditions. Additionally, the short survey completion time may have been influenced by the selection of two simple visual examples. Focusing exclusively on climate data introduces another constraint. While the topic is timely and relevant, it may evoke emotionally charged or politically motivated audience reactions, which may have influenced perception. Though we addressed topic-related responses in our qualitative analysis (see Section 4.1.5) and we asked respondents to self-assess their interest in the topic, we recognize that unmeasured biases may remain. Yet, studying perceptions of potentially polarizing visuals is key to understanding real-world information retrieval and addressing misinformation [90, 91].

5.3. Future Work

There are numerous promising directions for further developing the multidimensional VDL assessment proposed here, as well as several ideas for improving future survey conduction by varying test implementation conditions. Future implementations of MAVIL should be expanded to include a wider variety of visualizations formats. Beyond static charts, interactive visualizations are increasingly common in digital communication, possibly requiring different skills [92] and forms of audience engagement [93]. Testing MAVIL on such formats may warrant an additional dimension to assess interaction-based ability facets. Other common visualization types, such as infographics, could be explored to reflect broad visual data communication practices [94, 95, 96, 97].

MAVIL could further benefit from integration with existing tools like Mini-VLAT [9], as a combined assessment would allow for further comparison of self-perceived and task-based results within the same sample. Initial correlations we observed, such as between self-perceived numeracy and data reading performance, highlight promising ground for such work. Additionally, future work could draw on insights from recent assessment contributions, such as PREVis, which focuses on perceived readability in data visualizations [46]. PREVis scales could, for instance, enrichen MAVIL Dimension 2 on *Decoding* by incorporating validated survey questions from this contribution, such as those addressing perceived visualization layout understanding.

Though this study focused on climate data, MAVIL could potentially be adapted to other domains. Applying it to contexts like public health [98, 99] or social issues [17] could test its transferability and uncover specific challenges in audience comprehension and perception for other crucial domains. Adapting MAVIL questions to visualization content and topics has proven effective, such as in *Contextualizing*, where the *Digitize!* survey featured an expanded set of questions on climate-related interest and attitudes (refer to Figure 3.3). Although MAVIL was applied to climate data visualizations, its structure could suit other contexts. Testing varied MAVIL structures, including dimension arrangements, customized questions, and different visualization topics, aligns with research emphasizing that diverse users with varying literacy and contextual knowledge require tailored approaches [83, 84]. Additionally, a more systematic investigation into sequencing effects, such as whether the order of question blocks influences responses, could improve the instrument's reliability.

Building on the curiosity expressed by survey pilot participants about whether their visualization understanding was "correct," particularly during *Reading* tasks, future MAVIL iterations could introduce an educative feedback function. This could include providing feedback to survey responses, such as revealing correct answers to task-based questions. If specific misconceptions or errors arise during visualization tasks, these could be explicitly addressed [8]. Such feedback could also include summaries of respondents' self-perceived strengths and weaknesses across ability dimensions at the end of the survey, so that they are encouraged to reflect upon their VDL and visualization understanding. This feature should not come across as a moral imperative but rather complement the assessment with tailored education [38], and may serve as an incentive for respondents.

Lastly, integrating more diverse respondent perspectives, particularly concerning the cultural and socio-economic backgrounds, would enhance the assessment development. The survey data sets evaluated here may be biased due to distinct geographical, cultural, and socioeconomic backgrounds from the participant groups. While test implementations described in this thesis considered gender balance and included younger and older demographics, diversity could be expanded to encompass varying educational backgrounds and the dynamics of rural and urban living contexts. Future surveys could also target non-Western European audiences and focus on regions where the education on Visual Data Literacy or the perception of climate change differs. Highly educated respondents were overrepresented in our testing, which is why reaching audiences with less exposure to statistics or self-reflection on skills is an important aim for future work. This leads back to one of the main aims for MAVIL being that it should suit general audiences that are not expected to hold any expertise related to the surveyed visual representations.

Survey designs addressing these factors might deepen understanding of how context influences Visual Data Literacy. Including questions about respondents' lived experiences and opinions could also shed light on polarized perspectives on climate change, fostering a more inclusive discourse. Public awareness and attitudes toward global environmental change vary across countries, affecting visualization interpretation [37]. Broadening respondent diversity could ensure MAVIL remains accessible, easy to comprehend, and capable of supporting diverse formats for expressing visualization understanding.

5.4. Conclusion

MAVIL was developed as a novel approach to assessing Visual Data Literacy by conceptualizing it as a multidimensional ability composed of distinct facets. Grounded in a theory-driven approach, it introduces six ability dimensions and combines task-based, self-assessed, and open-ended question formats. This expands existing literacy assessments by capturing not only performance but also perception and qualitative input. Through survey pilots, a representative survey in Austria, and integration into the *Digitize!* project, MAVIL provided nuanced insights into people's (self-perceived) Visual Data Literacy and their understanding of two climate data visualizations. The findings revealed some deficits, including perceived challenges in interpreting simple measurement units and some unfamiliarity with presumably simple chart types. Overall, MAVIL assesses a combination of different yet central aspects of visualization understanding and aims to encourage more nuanced and critical approaches to evaluating Visual Data Literacy in general audiences.

Figure Credits for Surveyed Climate Data Visualizations

- Figure 1.1a: Statista, *Der Meeresspiegel steigt kontinuierlich an*, NASA, `https://de.statista.com/infografik/21922/anstieg-des-meeresspiegels/`
- Figure 1.1b: Statista, *Diese Länder stoßen am meisten CO_2 aus*, Global Carbon Project, `https://de.statista.com/infografik/18287/co2-emissionen-in-ausgewaehlten-laendern/`

Bibliography

[1] P. W. Singer and E. T. Brooking, *LikeWar: The weaponization of social media.* New York, NY, USA: Eamon Dolan Books, 2018. ISBN: 9780358108474.

[2] B. R. Sood, G. Stockdale, and E. M. Rogers, "How the news media operate in natural disasters," *Journal of communication*, vol. 37, no. 3, pp. 27–41, 1987. https://doi.org/10.1111/j.1460-2466.1987.tb00992.x.

[3] C. V. Bell and R. M. Entman, "The media's role in America's exceptional politics of inequality: Framing the Bush tax cuts of 2001 and 2003," *The International Journal of Press/Politics*, vol. 16, no. 4, pp. 548–572, 2011. https://doi.org/10.1177/1940161211417334.

[4] R. I. Field, "What you see is what you fear," *Human Vaccines & Immunotherapeutics*, vol. 9, no. 12, pp. 2670–2671, 2013. https://doi.org/10.4161/hv.26653.

[5] A. R. Tait, T. Voepel-Lewis, B. J. Zikmund-Fisher, and A. Fagerlin, "Presenting research risks and benefits to parents: does format matter?," *Anesthesia and analgesia*, vol. 111, no. 3, p. 718, 2010.

[6] S. T. Hawley, B. Zikmund-Fisher, P. Ubel, A. Jancovic, T. Lucas, and A. Fagerlin, "The impact of the format of graphical presentation on health-related knowledge and treatment choices," *Patient Education and Counseling*, vol. 73, no. 3, pp. 448–455, 2008. https://doi.org/10.1016/j.pec.2008.07.023.

[7] M. Chen, A. Abdul-Rahman, D. Archambault, J. Dykes, A. Slingsby, P. D. Ritsos, T. Torsney-Weir, C. Turkay, B. Bach, A. Brett, H. Fang, R. Jianu, S. Khan, R. S. Laramee, P. H. Nguyen, R. Reeve, J. C. Roberts, F. Vidal, Q. Wang, J. Wood, and K. Xu, "RAMPVIS: Towards a New Methodology for Developing Visualisation Capabilities for Large-scale Emergency Responses," Dec. 2020. https://doi.org/10.48550/arXiv.2012.04757.

[8] K. Sung-Hee, S. Lee, and B. C. Kwon, "VLAT: Development of a visualization literacy assessment test," *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 1, pp. 551–560, 2017. https://doi.org/10.1109/TVCG.2016.2598920.

[9] S. Pandey and A. Ottley, "Mini-vlat: A short and effective measure of visualization literacy," *Computer Graphics Forum*, vol. 42, no. 3, pp. 1–11, 2023. https://doi.org/10.1111/cgf.14809.

[10] L. W. Ge, Y. Cui, and M. Kay, "Calvi: Critical thinking assessment for literacy in visualizations," in *Conference on Human Factors in Computing Systems CHI*, (New York, New York, USA), pp. 815:1–815:18, Association for Computing Machinery, 2023. `https://doi.org/10.1145/3544548.3581406`.

[11] H. Fischer, S. Schütte, A. Depoux, D. Amelung, and R. Sauerborn, "How well do cop22 attendees understand graphs on climate change health impacts from the fifth ipcc assessment report?," *International Journal of Environmental Research and Public Health*, vol. 15, no. 5, p. 875, 2018. `https://doi.org/10.3390/ijerph15050875`.

[12] L. Chujun and M. A. Thornton, "Visualization aesthetics bias trust in science, news, and social media," *PsyArXiv*, 2021. `https://doi.org/10.31234/osf.io/dnr9s`.

[13] G. P. Wiggins and J. McTighe, *Understanding by Design.* Association for Supervision and Curriculum Development, expanded second edition ed., 2005. ISBN: 9780131950849.

[14] B. S. Bloom, *Taxonomy of Educational Objectives, Handbook: The Cognitive Domain.* New York, New York, USA: David McKay, 1956. ISBN: 9780582280106.

[15] S. C. Moser, "Communicating climate change: history, challenges, process and future directions," *Wiley Interdisciplinary Reviews: Climate Change*, vol. 1, no. 1, pp. 31–53, 2010. `https://doi.org/10.1002/wcc.11`.

[16] L. M. Padilla, S. H. Creem-Regehr, and W. Thompson, "The powerful influence of marks: Visual and knowledge-driven processing in hurricane track displays.," *Journal of experimental psychology: applied*, vol. 26, no. 1, p. 1, 2020. `https://doi.org/10.31234/osf.io/5tg9y`.

[17] P. Doboš, "Visualizing the european migrant crisis on social media: the relation of crisis visualities to migrant visibility," *Geografiska Annaler: Series B, Human Geography*, vol. 105, no. 1, pp. 99–115, 2023. `https://doi.org/10.1080/04353684.2022.2098156`.

[18] P. Du, D. Li, T. Liu, L. Zhang, X. Yang, and Y. Li, "Crisis Map Design Considering Map Cognition," *ISPRS International Journal of Geo-Information*, vol. 10, p. 692, Oct. 2021.

[19] S. J. O'Neill and N. Smith, "Climate change and visual imagery," *Wiley Interdisciplinary Reviews: Climate Change*, vol. 5, no. 1, pp. 73–87, 2014. `https://doi.org/10.1002/wcc.249`.

[20] B. Schneider, "Climate model simulation visualization from a visual studies perspective," *Wiley Interdisciplinary Reviews: Climate Change*, vol. 3, no. 2, pp. 185–193, 2012. `https://doi.org/10.1002/wcc.162`.

[21] J. Lynn, *IPCC expert meeting on communication meeting report.* World Meteorological Organization, Geneva, 2016.

[22] J. Harold, I. Lorenzoni, T. Shipley, and K. Coventry, "Communication of IPCC visuals: IPCC authors' views and assessments of visual complexity," *Climate Change*, vol. 158, no. 2, pp. 255–270, 2019. `https://doi.org/10.1007/s10584-019-02537-z`.

[23] A. Thorpe, A. Scherer, P. J. K. Han, N. Burpo, V. Shaffer, L. Scherer, and A. Fagerlin, "Exposure to common geographic covid-19 prevalence maps and public knowledge, risk perceptions, and behavioral intentions," *JAMA Network Open*, vol. 4, 2021. `https://doi.org/10.1001/jamanetworkopen.2020.33538`.

[24] S. L. Franconeri, L. M. Padilla, P. Shah, J. M. Zacks, and J. Hullman, "The Science of Visual Data Communication: What Works," *Psychological Science in the Public Interest*, vol. 22, pp. 110–161, Dec. 2021. `https://doi.org/10.1177/15291006211051956`.

[25] J. Callow, "Show me: Principles for assessing students' visual literacy," *The Reading Teacher*, vol. 61, no. 8, pp. 616–626, 2008. `https://doi.org/10.1598/RT.61.8.3`.

[26] E. E. Firat, C. Lang, B. Srinivas, I. Peng, R. S. Laramee, and A. Joshi, "A Constructivism-based Approach to Treemap Literacy in the Classroom," in *Eurographics 2023 - Education Papers* (A. Magana and J. Zara, eds.), The Eurographics Association, 2023. `https://doi.org/10.2312/eged.20231016`.

[27] B. Lee, E. K. Choe, P. Isenberg, K. Marriott, and J. Stasko, "Reaching broader audiences with data visualization," *IEEE Computer Graphics and Applications*, vol. 40, no. 2, pp. 82–90, 2020. `https://doi.org/10.1109/MCG.2020.2968244`.

[28] Y. V. Vaerenbergh and T. Thomas, "Response styles in survey research: A literature review of antecedents, consequences, and remedies," *International Journal of Public Opinion Research*, vol. 25, pp. 195–217, 06 2013. `https://doi.org/10.1093/ijpor/eds021`.

[29] R. Schuster, K. Gregory, T. Möller, and L. Koesten, ""being simple on complex issues" – accounts on visual data communication about climate change," *IEEE transactions on visualization and computer graphics*, vol. PP, 01 2024. `https://doi.org/10.1109/TVCG.2024.3352282`.

[30] A. Michelson, "A short history of visual literacy: The first five decades," *Art Libraries Journal*, vol. 42, no. 2, pp. 95–98, 2017. `https://doi.org/10.1017/alj.2017.10`.

[31] D. Hattwig, "Acrl visual literacy competency standards for higher education," tech. rep., Association of College and Research Libraries, 2011. `https://ala.org/acrl/standards/visualliteracy`.

[32] K. Börner, A. Bueckle, and M. Ginda, "Data visualization literacy," *Proceedings of the National Academy of Sciences*, vol. 116, no. 6, pp. 1857–1864, 2019. `https://doi.org/10.1073/pnas.1807180116`.

[33] T. Bowen, "Assessing visual literacy: a case study of developing a rubric for identifying and applying criteria to undergraduate student learning," *Teaching in Higher Education*, vol. 22, no. 6, pp. 705–719, 2017. `https://doi.org/10.1080/13562517.2017.1289507`.

[34] R. S. Robinson, "Learning to see: Developing visual literacy through film," *Top of the News*, vol. 40, no. 3, pp. 267–275, 2017.

[35] T. He, P. Isenberg, R. Dachselt, and T. Isenberg, "BeauVis: A validated scale for measuring the aesthetic pleasure of visual representations," *IEEE Transactions on Visualization and Computer Graphics*, vol. 29, no. 1, pp. 363–373, 2023. `https://doi.org/10.1109/TVCG.2022.3209390`.

[36] V. F. Reyna, W. L. Nelson, P. K. Han, and N. F. Dieckmann, "How numeracy influences risk comprehension and medical decision making," *Psychol Bull*, vol. 135, no. 6, pp. 943–973, 2009. `https://doi.org/10.1037/a0017327`.

[37] G. Xexakis and E. Trutnevyte, "Empirical testing of the visualizations of climate change mitigation scenarios with citizens: A comparison among germany, poland, and france," *Global Environmental Change*, vol. 70, p. 102324, 2017. `https://doi.org/10.1016/j.gloenvcha.2021.102324`.

[38] F. Chevalier, N. Henry Riche, B. Alper, C. Plaisant, J. Boy, and N. Elmqvist, "Observations and reflections on visualization literacy in elementary school," *IEEE Computer Graphics and Applications*, vol. 38, no. 3, pp. 21–29, 2018. `https://doi.org/10.1109/MCG.2018.032421650`.

[39] J. Brill and R. Maribe Branch, "Visual literacy defined - the results of a delphi study: Can ivla (operationally) define visual literacy?," *Journal of Visual Literacy*, vol. 27, no. 1, pp. 47–60, 2018. `https://doi.org/10.1080/23796529.2007.11674645`.

[40] M. Hegarty, "Multimedia learning about physical systems," in *The Cambridge Handbook of Multimedia Learning (Cambridge Handbooks in Psychology* (R. Mayer, ed.), p. 447–465, Cambridge: Cambridge University Press, 2005. `https://doi.org/CBO9780511816819.029`.

[41] E. Lee-Robbins and E. Adar, "Affective learning objectives for communicative visualizations," *IEEE Transactions on Visualization and Computer Graphics*, vol. 29, no. 1, pp. 1–11, 2023. `https://doi.org/10.1109/TVCG.2022.3209500`.

[42] E. E. Firat, A. Denisova, and R. S. Laramee, "Treemap literacy: A classroom-based investigation," in *Eurographics 2020 - Education Papers* (M. Romero and B. Sousa Santos, eds.), The Eurographics Association, 2020. `https://doi.org/10.2312/eged.20201032`.

[43] E. E. Firat, A. Denisova, M. L. Wilson, and R. S. Laramee, "P-lite: A study of parallel coordinate plot literacy," *Visual Informatics*, vol. 6, no. 3, pp. 81–99, 2022. `https://doi.org/10.1016/j.visinf.2022.05.002`.

[44] Y. Rogers and P. Marshall, *Research in the Wild*. Springer Cham, 2017. https://doi.org/10.1007/978-3-031-02220-3.

[45] 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," *Proceedings of the National Academy of Sciences*, vol. 15, no. 3, pp. 198–213, 2016. https://doi.org/10.1177/1473871615594652.

[46] A.-F. Cabouat, T. He, P. Isenberg, and T. Isenberg, "Previs: Perceived readability evaluation for visualizations," *IEEE Transactions on Visualization and Computer Graphics*, vol. 31, p. 1083–1093, Sept. 2024. https://doi.org/10.1109/TVCG.2024.3456318.

[47] I. Stuart-Hamilton, *Dictionary of psychological testing, assessment and treatment*. Jessica Kingsley Publishers, 2 ed., 2007. ISBN: 1843104946.

[48] J. Allen and R. Velden, *The role of self-assessment in measuring skills*. REFLEX, 2005. "https://cris.maastrichtuniversity.nl/en/publications/the-role-of-self-assessment-in-measuring-skills" (accessed 2025-05-14).

[49] R. Garcia-Retamero, E. T. Cokely, S. Ghazal, and A. Joeris, "Measuring graph literacy without a test," *Medical Decision Making*, vol. 36, no. 7, p. 854–867, 2016. https://doi.org/10.1177/0272989X16655334.

[50] C. A. Bauer, V. Job, and B. Hannover, "Who gets to see themselves as talented? biased self-concepts contribute to first-generation students' disadvantage in talent-focused environments," *Journal of Experimental Social Psychology*, vol. 108, p. 104501, 2023. https://doi.org/10.1016/j.jesp.2023.104501.

[51] B. Spinath, F. M. Spinath, N. Harlaar, and R. Plomin, "Predicting school achievement from general cognitive ability, self-perceived ability, and intrinsic value," *Intelligence*, vol. 34, no. 4, pp. 363–374, 2006. https://doi.org/10.1016/j.intell.2005.11.004.

[52] C. Greven, N. Harlaar, Y. Kovas, T. Chamorro-Premuzic, and R. Plomin, "More than just iq school achievement is predicted by self-perceived abilities—but for genetic rather than environmental reasons," *Psychological science*, vol. 20, pp. 753–62, 06 2009. https://doi.org/10.1111/j.1467-9280.2009.02366.x.

[53] E. Nohlert, J. Ohrvik, and A. Helgason, "Self-perceived ability to cope with stress and depressive mood without smoking predicts successful smoking cessation 12 months later in a quitline setting: A secondary analysis of a randomized trial," *BMC Public Health*, vol. 18, 08 2018. https://doi.org/10.1186/s12889-018-5973-9.

[54] D. Morisi, J. T. Jost, and V. Singh, "An asymmetrical "president-in-power" effect," *American Political Science Review*, vol. 113, no. 2, p. 614–620, 2019. https://doi.org/10.1017/S0003055418000850.

[55] S. Kritzinger, M. Foucault, R. Lachat, J. Partheymüller, C. Plescia, and S. B. and, "'rally round the flag': the covid-19 crisis and trust in the national government," *West European Politics*, vol. 44, no. 5-6, pp. 1205–1231, 2021. https://doi.org/10.1080/01402382.2021.1925017.

[56] S. Alturki and H. Stuckenschmidt, "Assessing students' self-assessment ability in an interdisciplinary domain," *Journal of applied research in higher education*, vol. 14, no. 4, pp. 1449–1465, 2022. https://doi.org/10.1108/JARHE-01-2021-0034.

[57] M. Davis and D. Lohm, *Pandemics, Publics, and Narrative*. Oxford, UK: Oxford University Press, 05 2020. https://doi.org/10.1093/oso/9780190683764.001.0001.

[58] D. R. Garfin, R. C. Silver, and E. A. Holman, "The novel coronavirus (COVID-2019) outbreak: Amplification of public health consequences by media exposure," *Health psychology*, vol. 39, no. 5, pp. 355–357, 2020. https://doi.org/10.1037/hea0000875.

[59] Y. Okan, R. Garcia-Retamero, E. T. Cokely, and A. Maldonado, "Individual differences in graph literacy: Overcoming denominator neglect in risk comprehension," *Journal of Behavioral Decision Making*, vol. 25, no. 4, pp. 390–401, 2012. https://doi.org/doi.org/10.1002/bdm.751.

[60] L. W. Ge, Y. Cui, and M. Kay, "Avec: An assessment of visual encoding ability in visualization construction," in *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, CHI '25, (New York, NY, USA), Association for Computing Machinery, 2025. https://doi.org/10.1145/3706598.3713364.

[61] L. M. Christian and D. A. Dillman, "The influence of graphical and symbolic language manipulations on responses to self-administered questions," *Public Opinion Quarterly*, vol. 68, no. 1, p. 57–80, 2004. https://doi.org/10.1093/poq/nfh004.

[62] F. Kreuter, S. Presser, and R. Tourangeau, "Social desirability bias in cati, ivr, and web surveys: The effects of mode and question sensitivity," *Public Opinion Quarterly*, vol. 72, no. 5, pp. 847–865, 2008. https://doi.org/10.1093/poq/nfn063.

[63] M. T. Boykoff, *Who Speaks for the Climate? Making Sense of Media Reporting on Climate Change*. Cambridge, UK; New York, USA: Cambridge University Press, 2011. https://doi.org/10.1017/CBO9780511978586.

[64] U. of Vienna, "Digitize! computational social sciences in der digitalen und sozialen transformation." "https://digitize.univie.ac.at/en/" (website, accessed 2024-03-25).

[65] D. J. Leiner, "Sosci survey." https://www.soscisurvey.de (accessed 2023-09-29).

[66] R. Tourangeau, L. J. Rips, and K. Rasinski, *Respondents' Understanding of Survey Questions*, p. 23–61. Cambridge University Press, 2000. https://doi.org/10.1017/CBO9780511819322.

[67] A. I. Elaine and C. A. Seaman, "Likert scales and data analyses," *Quality Progress*, vol. 40, 2007. "https://www.bayviewanalytics.com/reports/asq/likert-scales-and-data-analyses.pdf" (accessed 2025-05-14).

[68] L. Wilson and E. Dickinson, *Respondent Centred Surveys: Stop, Listen and Then Design.* Thousand Oaks: SAGE Publications Ltd., 1 ed., 2022. ISBN: 1529701260.

[69] J. E. A. McColl, L. Thomas, J. Soutter, B. C., N. Steen, R. Thomas, E. Harvey, A. Garratt, and J. Bond, "Design and use of questionnaires: A review of best practice applicable to surveys of health service staff and patients," *HTA on DVD/Health Technology Assessment*, vol. 31, no. 5, p. 1–256, 2001. https://doi.org/10.3310/hta5310.

[70] 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*, vol. 20, no. 12, pp. 1963–1972, 2017. https://doi.org/10.1109/TVCG.2014.2346984.

[71] J. M. Corbin and A. L. Strauss, *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory.* Los Angeles/London: Sage Publications, 4 ed., 2015. https://doi.org/10.4135/9781452230153.

[72] A. Bryant, *The Sage Handbook of Grounded Theory.* Los Angeles/London: Sage Publications, 2010. https://doi.org/10.4135/9781848607941.

[73] S. Austria, "Bildungsstand der Bevölkerung kernel description," 2023. "https://www.statistik.at/statistiken/bevoelkerung-und-soziales/bildung/bildungsstand-der-bevoelkerung" (accessed 2023-07-07).

[74] S. Kritzinger, K. Pfaff, J. Barta, J. Bernhard, H. Boomgaarden, A. Eder, N. Forgó, F. Paspalj, C. Plant, B. Prainsack, D. Prandner, S. Rittel, M. Teuffenbach, and S. Tschiatschek, "Digitize! – computational social science in der digitalen und sozialen transformation," *Zeitschrift Für Hochschulentwicklung*, vol. 18, no. Sonderheft Forschung, pp. 173–195, 2023. https://doi.org/10.21240/zfhe/SH-F/11.

[75] U. of Vienna, "Methods and data. digitize! computational social sciences in der digitalen und sozialen transformation." "https://digitize.univie.ac.at/en/surveys/methods-and-data/" (accessed 2025-01-13).

[76] "Pearson's chi-squared test for count data (r documentation)." https://stat.ethz.ch/R-manual/R-patched/library/stats/html/chisq.test.html.

[77] J. Shijie, P. Yi, L. Weisheng, and S. Liyin, "Study on analyzing questionnaire survey by monte carlo simulation," in *2010 International Conference on E-Business and E-Government*, pp. 1452–1455, 2010. https://doi.org/10.1109/ICEE.2010.369.

[78] S. Demarest, J. Van der Heyden, R. Charafeddine, J. Tafforeau, H. Van Oyen, and G. Van Hal, "Socio-economic differences in participation of households in a belgian national health survey," *European Journal of Public Health*, vol. 23, no. 6, pp. 981–985, 2013.

[79] H. Kennedy and R. L. Hill, "The feeling of numbers," *Sociology (Oxford)*, vol. 52, no. 4, pp. 830–848, 2018. https://doi.org/10.1177/0038038516674675.

[80] L. Koesten, T. Möller, and A. Saske, "Digitize! news (05/2024)," 2024. "https://digitize.univie.ac.at/fileadmin/user_upload/p_digitize/Kurzbericht_W5.pdf" (accessed 2025-05-14).

[81] L. Koesten, A. Saske, S. Starchenko, and K. Gregory, "Encountering friction, understanding crises: How do digital natives make sense of crisis maps?," in *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, CHI '25, (New York, NY, USA), Association for Computing Machinery, 2025. https://doi.org/10.1145/3706598.3713520.

[82] J. Kolko, "Sensemaking and framing: A theoretical reflection on perspective in design synthesis," in *Design and Complexity - DRS International Conference 2010* (D. Durling, R. Bousbaci, L. Chen, P. Gauthier, T. Poldma, S. Roworth-Stokes, and E. Stolterman, eds.), (Montreal, Canada), July 7–9 2010. "https://dl.designresearchsociety.org/drs-conference-papers/drs2010/researchpapers/67" (accessed 2025-05-14).

[83] J. Kostelnick, D. Mcdermott, R. Rowley, and N. Bunnyfield, "A Cartographic Framework for Visualizing Risk," *Cartographica: The International Journal for Geographic Information and Geovisualization*, vol. 48, pp. 200–224, Sept. 2013. https://doi.org/10.3138/carto.48.3.1531.

[84] C. S. Skinner, V. Strecher, and H. Hospers, "Physicians' recommendations for mammography: Do tailored messages make a difference?," *American Journal of Public Health*, vol. 84, no. 1, pp. 43–49, 1994. https://doi.org/10.2105/AJPH.84.1.43.

[85] J. D. Camba, P. Company, V. Byrd, B. Santos, and G. Alford, "Identifying deception as a critical component of visualization literacy," *IEEE Computer Graphics and Applications*, vol. 42, pp. 116–122, 01 2022. https://doi.org/10.1109/MCG.2021.3132004.

[86] M. Dörk, P. Feng, C. Collins, and S. Carpendale, "Critical infovis: exploring the politics of visualization," in *CHI '13 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '13, (New York, NY, USA), p. 2189–2198, Association for Computing Machinery, 2013. https://doi.org/10.1145/2468356.2468739.

[87] A. Hajek and H.-H. Konig, "Level and correlates of empathy and altruism during the covid-19 pandemic. evidence from a representative survey in germany," *PloS One*, vol. 17, no. 3, 2022. `https://doi.org/10.1371/journal.pone.0265544`.

[88] S. C. Cates, K. M. Kosa, J. E. Brophy, A. J. Hall, and A. Fraser, "Consumer education needed on norovirus prevention and control: Findings from a nationally representative survey of u.s. adults," *Journal of Food Protection*, vol. 78, no. 3, pp. 484–490, 2015. `https://doi.org/10.4315/0362-028X.JFP-14-313`.

[89] N. Mattheos, A. Nattestad, E. Falk-Nilsson, and R. Attström, "The interactive examination: assessing students' self-assessment ability," *Medical education*, vol. 38, no. 4, pp. 378–389, 2004. `https://doi.org/10.1046/j.1365-2923.2004.01788.x`.

[90] M. Lisnic, C. Polychronis, A. Lex, and M. Kogan, "Misleading beyond visual tricks: How people actually lie with charts," in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23, (New York, NY, USA), ACM, 2023. `https://doi.org/10.1145/3544548.3580910`.

[91] E. Ferrara, S. Cresci, and L. Luceri, "Misinformation, manipulation, and abuse on social media in the era of COVID-19," *Journal of Computational Social Science*, vol. 3, pp. 271–277, 2020. `https://doi.org/10.1007/s42001-020-00094-5`.

[92] H. Jia and S. Sundar, "Vivid and engaging: Effects of interactive data visualization on perceptions and attitudes about social issues," *Digital Journalism*, vol. 12, no. 8, pp. 1205–1229, 2024. `https://doi.org/10.1080/21670811.2023.2250815`.

[93] E. E. Firat, A. Joshi, and R. S. Laramee, "Interactive visualization literacy: The state-of-the-art," *Information Visualization*, vol. 21, no. 3, pp. 285–310, 2022. `https://doi.org/10.1177/14738716221081831`.

[94] A. Lazard and L. Atkinson, "Putting environmental infographics center stage," *Science Communication*, vol. 37, no. 1, pp. 6–33, 2017. `https://doi.org/10.1177/1075547014555997`.

[95] G. Huang, K. Li, and H. Li, "Show, not tell: The contingency role of infographics versus text in the differential effects of message strategies on optimistic bias," *Science Communication*, vol. 41, no. 6, pp. 732–760, 2019. `https://doi.org/10.1177/1075547019888659`.

[96] G. D. Haddow and K. Haddow, *Disaster Communications in a Changing Media World*. Amsterdam: Butterworth-Heinemann, 2 ed., 2014. `https://doi.org/10.1016/C2012-0-06592-1`.

[97] U. Hahn and P. Berkers, "Visualizing climate change: An exploratory study of the effectiveness of artistic information visualizations," *World Art*, vol. 11, no. 1, pp. 95–119, 2021. `https://doi.org/10.1080/21500894.2020.1769718`.

[98] M. Fan, Y. Wang, Y. Xie, F. M. Li, and C. C. and, "Understanding how older adults comprehend covid-19 interactive visualizations via think-aloud protocol," *International Journal of Human–Computer Interaction*, vol. 39, no. 8, pp. 1626–1642, 2023. `https://doi.org/10.1080/10447318.2022.2064609`.

[99] C. Juergens, "Trustworthy COVID-19 Mapping: Geo-spatial Data Literacy Aspects of Choropleth Maps," *KN - Journal of Cartography and Geographic Information*, vol. 70, no. 4, pp. 155–161, 2020. `https://doi.org/10.1007/s42489-020-00057-w`.

A. Appendix

MAVIL Catalog of Questions (Online Survey, Translation)

On **pages 58 through 64** is an English translation of the catalog of MAVIL questions. This catalog was implemented during an online survey in March 2023 (described in Section 3.3) with 438 respondents, representative of Austria (ages 18-74, male-female gender split). For the original German language version of the MAVIL catalog of questions used in the implementation, please refer to **pages 66 through 72**.

MAVIL CATALOG OF QUESTIONS
(Multidimensional Assessment of Visual Data Literacy)

Please note that MAVIL was designed in German, the primary language of the surveyed country, and also tested in this language version. The assessment questions were translated into English for this thesis, without having tested this language adaption.

INTRODUCTION

Dear participant,

with this questionnaire, you are taking part in a research project of the Digital Humanities Master's program at the University of Vienna. The project deals with the **understanding of data visualizations**. In this questionnaire, your personal experience while dealing with two visualizations is of interest.

No prior knowledge is required. The questionnaire part for each of the two visualizations can be completed in about 5 minutes. Please make sure you answer every question. Exactly one answer can be selected for each question, unless there is an option for open answers in text form.

This project was funded by the Vienna Science and Technology Fund (WWTF) and the City of Vienna through the ICT20-065 project. Your questionnaire will not be passed on to third parties.

Thank you for your participation!

Continue..1 == Continue

QC (CONSENT)

QCa I hereby confirm that I have read and understood this text and that I am over 16 years of age.

Consent given...1 == Start of survey

QD (DEMOGRAPHICS)

Please fill in the following details.

QDa Age in years

Below 18..1 == End of survey
18 to 24..2
25 to 34..3
35 to 44..4
45 to 54..5
55 to 64..6
65 to 74..7
75 or above...8

QDb Gender

Male..1
Female..2
Diverse..3
Inter..4
Open...5
Without gender entry...6
I do not wish to make a specification here..7

QDc Is your place of residence in Austria?

Yes.....1
No.....2 == End of survey

QDd Highest education level

No compulsory education.....1
Compulsory school certificate.....2
Baccalaureate.....3
University of applied sciences.....4
University degree.....6
Apprenticeship or other specialized training.....7
Master craftsman degree.....8
My degree is not listed. 9

INTRO II Climate change is shaping our current society and raises many questions about the future.

Data visualizations are often used to provide information about climate change. In this questionnaire you will see two visualizations on fact-based aspects of climate change.

Please look at the two visualizations and pay attention to both their content and your overall impression of the figure.

Continue.....1 == Continue

APPLICATION OF Q1-Q5: DATA VISUALIZATION 1 (V1)

Please look at the visualization and continue when you have grasped the message of the figure.
You will still be able to see the illustration on the following pages of the questionnaire.
You can enlarge or reduce the visualization by zooming in or out.
[show V1]

Continue.....1 == Continue

Q1-V1[1] Please assess the extent to which the statements apply.
The visualization can be found at the bottom of the page.
[show V1 beneath items]

Q1a-V1 The visualization seems understandable to me. The information can be clearly read.
Q1b-V1 I can classify the information presented in my previous knowledge and make connections.
Q1c-V1 I could summarize the content of the visualization for someone.

I disagree.....1
I somewhat disagree.....2
Undecided.....3
I somewhat agree.....4
I agree.....5

[1] V1 refers to visualization example 1. Please refer to the catalog’s last page for the surveyed data visualizations.

Q2-V1 Please assess the extent to which the statements apply.
[show V1 beneath items]

Q2a-V1 I can understand the units of measurement shown here, e.g., millimeters.
Q2b-V1 I can easily see whether the colors used in the visualization show any specific information.
Q2c-V1 I am familiar with the line chart and it is not the first time that I have read information from such a chart.

I disagree......1
I somewhat disagree......2
Undecided......3
I somewhat agree......4
I agree......5

Q3-V1 Please assess the extent to which the statements apply.
[show V1 beneath items]

Q3a-V1 This visualization appeals to me.
Q3b-V1 This visualization is pretty to look at.
Q3c-V1 This visualization is well designed.
Q3d-V1 This visualization is trustworthy.

I disagree......1
I somewhat disagree......2
Undecided......3
I somewhat agree......4
I agree......5

Q4-V1 Please assess the extent to which the statements apply.
[show V1 beneath items]

Q4a-V1 I have found an error in the visualization.
Q4b-V1 I have a suggestion for a change to the visualization.
Q4c-V1 If I could, I would design something in the visualization differently.

I disagree......1
I somewhat disagree......2
Undecided......3
I somewhat agree......4
I agree......5

Q4d-V1 I would like to remark something positive about the visualization:
Q4e-V1 I would like to remark something negative about the visualization:

Text field (max. 2000 characters, optional) ......TEXT

Q5-V1 Please decide whether the statements are true or false according to what the visualization shows.
[show V2 beneath items]

Q5a-V1 The sea level fluctuates, showing no trend.
True..0 == Incorrect
False..1 == Correct

Q5b-V1 In 2021, the sea level has risen by around ten centimeters compared to 1993.
True..1 == Correct
False..0 == Incorrect

APPLICATION OF Q1-Q5: DATA VISUALIZATION 2 (V2)

Please look at the visualization and continue when you have grasped the message of the figure.
You will still be able to see the illustration on the following pages of the questionnaire.
You can enlarge or reduce the visualization by zooming in or out.
[show V2]

Continue..1 == Continue

Q1-V2 Please assess the extent to which the statements apply.
The visualization can be found at the bottom of the page.
[show V2 beneath items]

Q1a-V2 The visualization seems understandable to me. The information can be clearly read.
Q1b-V2 I can classify the information presented in my previous knowledge and make connections.
Q1c-V2 I could summarize the content of the visualization for someone.

I disagree..1
I somewhat disagree..2
Undecided..3
I somewhat agree..4
I agree..5

Q2-V2 Please assess the extent to which the statements apply.
[show V2 beneath items]

Q2a-V2 I can understand the units of measurement shown here, e.g., millions of tons.
Q2b-V2 I can easily see whether the colors used in the visualization show any specific information.
Q2c-V2 I am familiar with the bar chart and it is not the first time that I have read information from such a chart.
I disagree..1
I somewhat disagree..2
Undecided..3
I somewhat agree..4
I agree..5

Q3-V2 Please assess the extent to which the statements apply.
[show V2 beneath items]

Q3a-V2 This visualization appeals to me.
Q3b-V2 This visualization is pretty to look at.
Q3c-V2 This visualization is well designed.
Q3d-V2 This visualization is trustworthy.

I disagree..1
I somewhat disagree...2
Undecided..3
I somewhat agree..4
I agree..5

Q4-V2 Please assess the extent to which the statements apply.
[show V2 beneath items]

Q4a-V2 I have found an error in the visualization.
Q4b-V2 I have a suggestion for a change to the visualization.
Q4c-V2 If I could, I would design something in the visualization differently.

I disagree..1
I somewhat disagree...2
Undecided..3
I somewhat agree..4
I agree..5

Q4d-V2 I would like to remark something positive about the visualization:
Q4e-V2 I would like to remark something negative about the visualization:

Text field (max. 2000 characters, optional) ....TEXT

Q5-V2 Please decide whether the statements are true or false according to what the visualization shows.
[show V2 beneath items]

Q5a-V2 Only Russia and Germany show a decrease in CO_2 emissions over the 28 years shown.
True...1 == correct
False...0 == incorrect

Q5b-V2 Japan had the lowest increase in CO_2 emissions in 2018 compared to 1990.
True...0 == incorrect
False...1 == correct

Please answer the following question.

Q5c-V2 Which country is listed first in the visualization?

Text field (max. 2000 char., obligatory.........TEXT, „China“ (or similar) == correct

QD Please assess the extent to which the statements apply.

QDe I deal a lot with the topic of climate change out of my own interest.

QDf I have a good knowledge of mathematics, especially in the field of statistics.

I disagree.....1
I somewhat disagree.....2
Undecided.....3
I somewhat agree.....4
I agree.....5

QF (FEEDBACK)

QFa Any other feedback or comments you would like to make on this questionnaire:

Text field (max. 2000 characters, optional) .....TEXT

CLOSING Thank you for your participation!
Your answers have been saved. You can now close this browser window.

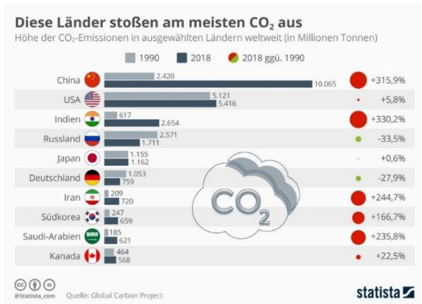
SURVEYED CLIMATE DATA VISUALIZATIONS

For a description of the data visualization's content and visual features in English please refer to Section 2.3 in the thesis.

V1: A line chart on sea level data.

(Source: Statista, *Der Meeresspiegel steigt kontinuierlich an*, NASA, https://de.statista.com/infografik/21922/anstieg-des-meeresspiegels/)

V2: A bar chart showing selected countries' CO_2 emissions over time.



(Source: Statista, *Diese Länder stoßen am meisten CO aus*, Global Carbon Project, URL: https://de.statista.com/infografik/18287/co2-emissionen-in-ausgewaehlten-laendern/)

MAVIL Fragenkatalog (Online Survey, Original)

On **pages 66 through 72** follows the German-language original version of the catalog of questions provided in the prior Section on **pages 58 through 64**.

MAVIL FRAGENKATALOG
(Multidimensional Assessment of Visual Data Literacy)

INTRODUKTION

Sehr geehrte*r Teilnehmer*in,

mit diesem Fragebogen nehmen Sie an einem Forschungsprojekt der Studienrichtung Digital Humanities der Universität Wien teil. Das Projekt setzt sich mit dem **Verstehen von Datenvisualisierungen** auseinander. In diesem Fragebogen ist Ihre persönliche Erfahrung während der Auseinandersetzung mit zwei Visualisierungen von Interesse.

Es sind keine Vorkenntnisse notwendig. **Der Fragebogenteil zu jeder der zwei Visualisierungen lässt sich in circa 5 Minuten abschließen.** Achten Sie bitte darauf, jede Frage zu beantworten. Pro Frage lässt sich genau eine Antwort auswählen, außer es besteht die Möglichkeit zur offenen Beantwortung in Textform.

Dieses Projekt wurde vom Wiener Wissenschafts- und Technologiefonds (WWTF) und der Stadt Wien durch das Projekt ICT20-065 finanziert. Ihr Fragebogen wird nicht an Dritte weitergegeben.

Vielen Dank für Ihre Teilnahme!

Weiter...1 == Weiter

QZ (ZUSTIMMUNG)

QZa Hiermit bestätige ich, dass ich diesen Text gelesen und verstanden habe und dass ich über 16 Jahre alt bin.

Zustimmung gegeben...1 == Fragebogenstart

QD (DEMOGRAPHIE)

Bitte füllen Sie folgenden Angaben aus.

QDa Alter in Jahren

Unter 18..1 == Fragebogenbeendigung
18 bis 24...2
25 bis 34...3
35 bis 44...4
45 bis 54...5
55 bis 64...6
65 bis 74...7
75 oder älter...8

QDb Geschlecht

Männlich...1
Weiblich...2
Divers...3
Inter..4
Offen...5
Ohne Geschlechtseintrag...6
Ich möchte hier keine Angabe tätigen. 7

QDc Liegt Ihr Wohnsitz in Österreich?

Ja..1
Nein...2 == Fragebogenbeendigung

QDd Abschlussgrad

Kein Pflichtschulabschluss..1
Pflichtschulabschluss...2
Matura…..3
Fachhochschulabschluss..4
Studienabschluss..6
Lehre oder eine andere Fachausbildung..7
Meisterabschluss...8
Mein Abschlussgrad ist nicht gelistet. 9

HINLEITUG Der Klimawandel prägt unsere gegenwärtige Gesellschaft und wirft viele zukunftsgebundene Fragen auf.

Um über den Klimawandel zu informieren, werden oft Datenvisualisierungen verwendet. In diesem Fragebogen sehen Sie zwei Visualisierungen zu faktenbasierten Aspekten des Klimawandels.

Bitte setzen Sie sich mit den beiden Visualisierungen auseinander und achten Sie sowohl auf deren Inhalt als auch auf Ihren Gesamteindruck der jeweiligen Abbildung.

Weiter...1 == Weiter

ANWENDUNG Q1-Q5: DATENVISUALISIERUNG 1 (V1)

Bitte betrachten Sie die Visualisierung und fahren Sie fort, wenn Sie die Aussage der Abbildung erfasst haben.
Sie werden die Abbildung auf den kommenden Seiten des Fragebogens weiterhin sehen können.
Sie können die Visualisierung durch Zoomen vergrößern oder verkleinern.
[Visualisierung einblenden]

Weiter...1 == Weiter

Q1-V1 Bitte schätzen Sie ein, inwiefern die Aussagen zutreffen.
Die Visualisierung befindet sich am Seitenende.
[Visualisierung unter Fragenblock einblenden]

Q1a-V1 Mir erscheint die Visualisierung verständlich. Die dargestellten Informationen lassen sich klar ablesen.
Q1b-V1 Ich kann die dargestellten Informationen in mein bisheriges Wissen einordnen und Zusammenhänge herstellen.
Q1c-V1 Ich könnte jemandem zusammenfassend den Inhalt der Visualisierung erzählen.

Trifft nicht zu..1
Trifft eher nicht zu..2
Teils-teils..3
Trifft eher zu..4
Trifft voll zu..5[1]

Q2-V1 Bitte schätzen Sie ein, inwiefern die Aussagen zutreffen.
[Visualisierung unter Fragenblock einblenden]

Q2a-V1 Ich kann die dargestellten Maßeinheiten, hier z.B. Millimeter, nachvollziehen.
Q2b-V1 Ich kann leicht erkennen, ob die verwendeten Farben in der Visualisierung bestimmte Informationen zeigen.
Q2c-V1 Das Liniendiagramm ist mir als Darstellungsform vertraut und es ist nicht das erste Mal, dass ich Informationen aus einem solchen Diagramm ablese.

Trifft nicht zu..1
Trifft eher nicht zu..2
Teils-teils..3
Trifft eher zu..4
Trifft voll zu..5

Q3-V1 Bitte schätzen Sie ein, inwiefern die Aussagen zutreffen.
[Visualisierung unter Fragenblock einblenden]

Q3a-V1 Die Visualisierung spricht mich an.
Q3b-V1 Die Visualisierung ist sehr schön anzusehen.
Q3c-V1 Die Visualisierung ist gut gestaltet.
Q3d-V1 Die Visualisierung scheint vertrauenswürdig.

Trifft nicht zu..1
Trifft eher nicht zu..2
Teils-teils..3
Trifft eher zu..4
Trifft voll zu..5

Q4-V1 Bitte schätzen Sie ein, inwiefern die Aussagen zutreffen.
[Visualisierung unter Fragenblock einblenden]

Q4a-V1 Ich habe einen Fehler in der Visualisierung gefunden.
Q4b-V1 Ich habe einen Änderungsvorschlag für die Visualisierung.
Q4c-V1 Wenn ich könnte, würde ich etwas in der Visualisierung anders gestalten.

Trifft nicht zu..1
Trifft eher nicht zu...2
Teils-teils..3
Trifft eher zu...4
Trifft voll zu...5

Q4d-V1 Ich möchte etwas Positives an der Visualisierung hervorheben:
Q4e-V1 Ich möchte etwas Negatives an der Visualisierung hervorheben:

Textfeld (max. 2000 Zeichen, optional) ....TEXT

Q5-V1 Bitte entscheiden Sie, ob die Aussagen laut der Visualisierung wahr oder falsch sind.
[Visualisierung unter Fragenblock einblenden]

Q5a-V1 Die Höhe des Meeresspiegels schwankt und zeigt keine Tendenz.
Q5b-V1 Im Jahr 2021 ist der Meeresspiegel im Vergleich zu 1993 um circa zehn Zentimeter gestiegen.

Wahr..1
Falsch..2

Bitte beantworten Sie die folgende Frage.

Q5c-V1 Wie hoch ist der Meeresspiegel in Millimetern im Jahr 2021?

Textfeld (max. 2000 Zeichen, obligatorisch) ....TEXT

ANWENDUNG Q1-Q5: DATENVISUALISIERUNG 2 (V2)

Bitte betrachten Sie die Visualisierung und fahren Sie fort, wenn Sie die Aussage der Abbildung erfasst haben.
Sie werden die Abbildung auf den kommenden Seiten des Fragebogens weiterhin sehen können.
Sie können die Visualisierung durch Zoomen vergrößern oder verkleinern.
[Visualisierung einblenden]

Weiter...1 == Weiter

Q1-V2 Bitte schätzen Sie ein, inwiefern die Aussagen zutreffen.
Die Visualisierung befindet sich am Seitenende.
[Visualisierung unter Fragenblock einblenden]

Q1a-V2 Mir erscheint die Visualisierung verständlich. Die dargestellten Informationen lassen sich klar ablesen.
Q1b-V2 Ich kann die dargestellten Informationen in mein bisheriges Wissen einordnen und Zusammenhänge herstellen.
Q1c-V2 Ich könnte jemandem zusammenfassend den Inhalt der Visualisierung erzählen.

Trifft nicht zu...1
Trifft eher nicht zu..2
Teils-teils..3
Trifft eher zu..4
Trifft voll zu...5

Q2-V1 Bitte schätzen Sie ein, inwiefern die Aussagen zutreffen.
[Visualisierung unter Fragenblock einblenden]

Q2a-V2 Ich kann die dargestellten Maßeinheiten, hier z.B. Tonnen, nachvollziehen.
Q2b-V2 Ich kann leicht erkennen, ob die verwendeten Farben in der Visualisierung bestimmte Informationen zeigen.
Q2c-V2 Das Balkendiagramm ist mir als Darstellungsform vertraut und es ist nicht das erste Mal, dass ich Informationen aus einem solchen Diagramm ablese.

Trifft nicht zu...1
Trifft eher nicht zu..2
Teils-teils..3
Trifft eher zu..4
Trifft voll zu...5

Q3-V2 Bitte schätzen Sie ein, inwiefern die Aussagen zutreffen.
[Visualisierung unter Fragenblock einblenden]

Q3a-V2 Die Visualisierung spricht mich an.
Q3b-V2 Die Visualisierung ist sehr schön anzusehen.
Q3c-V2 Die Visualisierung ist gut gestaltet.
Q3d-V2 Die Visualisierung scheint vertrauenswürdig.

Trifft nicht zu...1
Trifft eher nicht zu..2
Teils-teils..3
Trifft eher zu..4
Trifft voll zu...5

Q4-V2 Bitte schätzen Sie ein, inwiefern die Aussagen zutreffen.
[Visualisierung unter Fragenblock einblenden]

Q4a-V2 Ich habe einen Fehler in der Visualisierung gefunden.
Q4b-V2 Ich habe einen Änderungsvorschlag für die Visualisierung.
Q4c-V2 Wenn ich könnte, würde ich etwas in der Visualisierung anders gestalten.

Trifft nicht zu..1
Trifft eher nicht zu..2
Teils-teils...3
Trifft eher zu...4
Trifft voll zu..5

Q4d-V2 Ich möchte etwas Positives an der Visualisierung hervorheben:
Q4e-V2 Ich möchte etwas Negatives an der Visualisierung hervorheben:

Textfeld (max. 2000 Zeichen, optional) ....TEXT

Q5-V2 Bitte entscheiden Sie, ob die Aussagen laut der Visualisierung wahr oder falsch sind.
[Visualisierung unter Fragenblock einblenden]

Q5a-V2 Nur bei Russland und Deutschland zeigt sich ein Rückgang der CO_2-Emissionen in den dargestellten 28 Jahren.

Wahr...1 == korrekt
Falsch...0 == inkorrekt

Q5b-V2 Die geringste Zunahme an CO_2-Emissionen von 2018 gegenüber 1990 weist hier Japan vor.
Wahr...1 == korrekt
Falsch...0 == inkorrekt

Bitte beantworten Sie die folgende Frage.

Q5c-V2 Welches Landes wird in der Visualisierung zuerst aufgelistet?

Textfeld (max. 2000 Zeichen, obligatorisch)................TEXT, „China“ == korrekt

QD Bitte schätzen Sie ein, inwiefern die Aussagen zutreffen.
[Visualisierung unter Fragenblock einblenden]

QDe Ich setze mich aus eigenem Interesse viel mit dem Thema Klimawandel auseinander.
QDf Ich verfüge über gute Kenntnisse in der Mathematik, insbesondere im Bereich der Statistik.

Trifft nicht zu..1
Trifft eher nicht zu..2
Teils-teils...3
Trifft eher zu...4
Trifft voll zu..5

QF (Feedback)

QFa Sonstige Rückmeldungen oder Anmerkungen, die Sie zu diesem Fragebogen tätigen möchten:

Textfeld (max. 2000 Zeichen, optional) ....TEXT

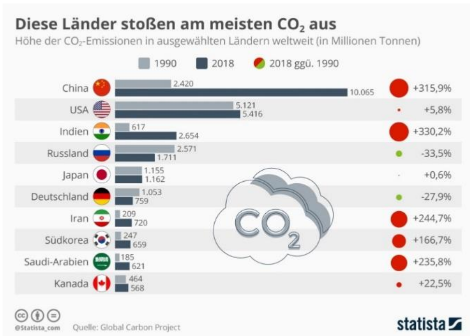
SCHLUSS Vielen Dank für Ihre Teilnahme!
Ihre Antworten wurden gespeichert, Sie können dieses Browser-Fenster nun schließen.

ANHANG: VERWENDETE VISUALISIERUNGSBEISPIELE

V1: Liniendiagramm

(Quelle: Statista, *Der Meeresspiegel steigt kontinuierlich an*, NASA, https://de.statista.com/infografik/21922/anstieg-des-meeresspiegels/)

V2: Balkendiagramm



(Quelle: Statista, *Diese Länder stoßen am meisten CO aus*, Global Carbon Project, URL: https://de.statista.com/infografik/18287/co2-emissionen-in-ausgewaehlten-laendern/)