

RESEARCH

Open Access



Strategies and difficulties during students' construction of data visualizations

Hsin-Yi Chang^{1*} , Yen-Jung Chang² and Meng-Jung Tsai^{1*} 

Abstract

Background Data visualizations transform data into visual representations such as graphs, diagrams, charts and so forth, and enable inquiries and decision-making in many professional fields, as well as in public and economic areas. How students' data visualization literacy (DVL), including constructing, comprehending, and utilizing adequate data visualizations, can be developed is gaining increasing attention in STEM education. As fundamental steps, the purpose of this study was to understand common student difficulties and useful strategies during the process of constructing data visualization so that suggestions and principles can be made for the design of curricula and interventions to develop students' DVL.

Methods This study engaged 57 college and high school students in constructing data visualizations relating to the topic of air quality for a decision-making task. The students' difficulties and strategies demonstrated during the process of data visualization were analyzed using multiple collected data sources including the students' think-aloud transcripts, retrospective interview transcripts, and process videos that captured their actions with the data visualization tool. Qualitative coding was conducted to identify the students' difficulties and strategies. Epistemic network analysis (ENA) was employed to generate network models revealing how the difficulties and strategies co-occurred, and how the college and high school students differed.

Results Six types of student difficulties and seven types of strategies were identified. The strategies were further categorized into non-, basic- and high-level metavisual strategies. About three-quarters of the participants employed basic or high-level metavisual strategies to overcome the technological and content difficulties. The high school students demonstrated a greater need to develop content knowledge and representation skills, whereas the college students needed more support to know how to simplify data to construct the best data visualizations.

Conclusions and implications The study specified metacognition needed for data visualization, which builds on and extends the cognitive model of drawing construction (CMDC) and theoretical perspectives of metavisualization. The results have implications for developing students' data visualization literacy in STEM education by considering the difficulties and trajectories of metacognitive strategy development, and by addressing the different patterns and needs demonstrated by the college and high school students.

Keywords Data visualization, Construction, Metacognitive strategy, Difficulty, Epistemic network analysis

*Correspondence:

Hsin-Yi Chang
hychang@ntnu.edu.tw
Meng-Jung Tsai
mjtsai99@ntnu.edu.tw

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Introduction

Data science is an emerging interdisciplinary field requiring the application of knowledge and skills in computer science, mathematics and statistics within specific domains such as science, economics or public health and policy (Ow-Yeong et al., 2023). Data visualization is a branch of data science that focuses on using visual representations such as graphs, charts or diagrams to represent data. The topic of data visualization is increasingly gaining attention in science, math and STEM education (Bybee, 2010; Donnelly-Hermosillo et al., 2020; Ow-Yeong et al., 2023).

In this digital age in which big data are available for scientific inquiry and analysis and decision-making, and with the rapid development of computer technology, data visualizations can be easily generated via computer visualization tools (Li, 2020; Unwin, 2020). Researchers suggest that it is important to develop all students' ability to construct, comprehend, and utilize adequate data visualizations, which comprises an important type of literacy called data visualization literacy [DVL] (Börner et al., 2016, 2019; Donohoe & Costello, 2020; Lee et al., 2017; Unwin, 2020). Development of learners' DVL facilitates a more data literate society (Bae et al., 2023). The current study addresses this call for all citizens to be data or information literate by focusing on how novices constructed data visualizations using a computer-based data visualization tool. The ability to construct data visualizations relates to a core practice of scientists and engineers working on "the construction of explanations or designs using reasoning, creative thinking, and models" (National Research Council [NRC], 2012, p. 44). Also, the use of modern technology is integral to the work. It is therefore important to provide novice students with opportunities to engage in such practice to develop their critical competencies to face challenges and issues in modern society (NRC, 2012).

Current research on DVL has mainly focused on developing and assessing students' ability to read, analyze and interpret data visualizations (Binali et al., 2022; Börner et al., 2019; Lee et al., 2017). However, an equally important issue with regard to how to develop students' ability to construct and utilize data visualizations has rarely been addressed. Indeed, a holistic data visualization literacy framework suggests not only interpretation, but also construction of data visualization as an important aspect of DVL (Börner et al., 2019).

As initial steps to develop students' ability to construct adequate data visualization, this current study engaged college and high school students in constructing data visualizations during individual interviews, and investigated their process of constructing data visualizations to identify the difficulties and strategies they demonstrated

during the process. Perspectives based on the constructivist theory indicate that students' ideas and their learning difficulties and strategies are building blocks for their learning. The development of instruction or interventions should take into account students' ideas and performances to address their needs (Linn et al., 2004).

Especially, learning strategies have been considered as a key aspect of academic performance in STEM education (Griese et al., 2015). Among all kinds of learning strategies such as cognitive and management strategies, metacognitive strategies play an important role in successful learning and problem solving. They involve higher-order skills including employment of metacognition such as planning, monitoring and reflecting (de Boer et al., 2018). Research has evidenced that students' use of metacognitive strategies can enhance their critical thinking (Ku & Ho, 2010), and problem solving (Blackford et al., 2023). However, research in data science education has yet to identify metacognitive strategies specific to the practice of data science and visualization. Such investigation would reveal important learning strategies that are keys to successful data visualization when students are learning data science.

This study applied the cognitive model of drawing construction (CMDC) (Van Meter & Firetto, 2013) and theoretical perspectives of metavisualization (Gilbert, 2005, 2008, 2010; Justi et al., 2009) and metarepresentational competence (diSessa, 2004) to investigate students' data visualization processes focusing on the demonstrated difficulties and strategies. Specifically, the construct of metavisual strategy has been proposed to refer to a type of metacognitive strategy for facilitating the process of visualization (Locatelli & Arroio, 2014, 2016; Locatelli & Davidowitz, 2021). Students' use of metacognitive strategies in the context of visualization, that is, their use of metavisual strategies, was specifically examined and reported. The following research questions (RQ) were addressed in this study:

RQ1: What are students' common difficulties during the process of constructing data visualizations?

RQ2: Whether and how do students generate and use metavisual strategies, as well as other strategies, to overcome the difficulties encountered during data visualization?

RQ3: Whether and how do students at two different educational levels, namely, college and high school, demonstrate different patterns of difficulties encountered and strategies used during data visualization?

Research has identified individual differences in students' employment of metacognitive strategies (Karlen et al., 2014). Another study found that college and high

school students performed significantly differently on items that measured their DVL in the aspects of comprehending and interpreting data visualizations (Binali et al., 2022). Therefore, this study explored how college and high school students differed in terms of the difficulties and strategies demonstrated when constructing data visualizations, to provide insights for the development of data science curricula that address the needs of learners at different educational levels. Moreover, this study employed epistemic network analysis (ENA) (Marquart et al., 2018; Shaffer et al., 2016; Wooldridge et al., 2018) which allows examination of the co-occurrence of the difficulties with the strategies, to better understand patterns of learners' construction of data visualization with regard to what difficulties can be overcome by what strategies.

Regarding advancement in theory, the results of this study provide empirical evidence for extension, elaboration and revision of the CMDC and metavisualization perspectives in the case of data visualization, which is a type of drawing or visualization with the aid of computer technology. As for implications for practice, the metavisual strategies identified in this study can serve as examples for curriculum designers to develop interventions that support novice learners to facilitate successful construction of data visualization and develop DVL, to address the call for more data science and data visualization curricula (Ow-Yeong et al., 2023).

Background

Research on data visualization in STEM education

The Pre-K-12 Guidelines for Assessment and Instruction in Statistics Education II (GAISE II), a Framework for Statistics and Data Science Education (Bargagliotti et al., 2020), delineate the goal of data science education as developing students' ability to formulate statistical investigative questions, collect and consider data, analyze data, and interpret results, enabling them to use data as problem solvers. Recent research has begun to propose methods for teaching data visualization in formal education (e.g., Byrd & Dwenger, 2021; Camm et al., 2023). For instance, a three-phase teaching process has been proposed, which includes creating the initial version of the visualization, sanitizing the data visualization, and refining data visualizations (Camm et al., 2023). Another study suggested the data visualization activity (DVA) worksheet method, which includes seven data visualization stages: acquire, parse, mine, filter and represent, critique, refine, and interact (Byrd & Dwenger, 2021). Research on teaching data visualization recognizes the higher-order thinking involved in teaching and learning data visualization (Byrd & Dwenger, 2021; Camm et al., 2023). However, these studies are perspective-based and did not investigate the effectiveness of the implementation or

identify students' difficulties associated with the higher-order thinking required when learning data visualization. Despite the importance of data visualization in professional fields and educational practices in the Big Data era, little research has investigated the process of students' data visualization to identify common difficulties and beneficial strategies so that remedies or interventions can be designed accordingly to promote students' DVL.

Data visualization may refer to both the process and the product. The process of data visualization involves cognitive activities such as inspecting datasets and transforming them into visualizations such as tables, graphs, diagrams or charts. The products of this process are data visualizations which represent data for data exploration, pattern identification, communication, and decision-making (Lee et al., 2017, 2019; Mansoor & Harrison, 2018). Börner et al. (2019) suggested a holistic data visualization literacy framework (DVL-FW) that incorporates different aspects related to DVL from multiple studies. The framework was established to facilitate the process of not only reading, but also constructing data visualizations. They suggested seven aspects that data visualization construction and interpretation need to consider, namely: (1) insight needs, also called basic task types (e.g., trends, correlations, comparisons); (2) types of data to be visualized (e.g., nominal, ordinal, ratio); (3) data analyses (e.g., statistical, temporal, topical); (4) visualizations (e.g., map, graph, chart); (5) graphic symbols (e.g., geometric, linguistic, or pictorial symbols); (6) graphic variables (e.g., position, color, motion); and (7) interactions (e.g., zoom, search and locate, filter). Moreover, they proposed a process of data visualization construction and interpretation, including five major steps, namely, acquiring relevant datasets and resources, analyzing data before they can be visualized, visualizing data by selecting and mapping between data and visualization types and symbols, deploying data visualizations for interactions, and interpreting the data visualization for real-world application (Börner et al., 2019).

The data visualization framework and process reveal the complexity that an individual needs to consider when constructing a data visualization. One most common difficulty of constructing a data visualization is inappropriate or poor choice of visualization type for a particular dataset, leading to the problem of the constructed visualization failing to convey clear meaning or not being aligned with the purpose (Chrysantina & Sæbø, 2019). More studies are needed to systematically investigate students' difficulties in creating data visualizations; the findings can then serve as building blocks for researchers to investigate ways to help students overcome the difficulties.

The majority of the research on data visualization in education has focused on how learners comprehend and interpret data visualizations (e.g., Binali et al., 2022; Börner et al., 2016; Lee et al., 2017). For example, research (e.g., Bertin, 2011; Lee et al., 2017; Wainer, 1992) has analyzed the knowledge and skills needed to adequately comprehend and interpret data visualizations, including knowledge of graph conventions, and the ability to depict and explain trends and relationships in graphs. It has also been noticed that learners' interpretation of data visualizations may be biased due to individual differences stemming from past experiences (Mansoor & Harrison, 2018).

Relatively few studies have focused on learners' construction of data visualizations. Among these studies, Grammel et al. (2010) investigated how novices constructed visualizations, and found a common difficulty whereby they often relied heavily on their prior experiences with data visualization types, and showed inconsistent use of data visualizations. Moreover, three types of challenges were identified, namely decomposing questions and goals into data attributes, designing visual mappings, and interpreting visualizations (Grammel et al., 2010). Drawing on previous research, this study further investigated what strategies were used by students to overcome the difficulties they confronted during the process of creating data visualizations.

Another study engaged children of age 6 to 11 years old in informal learning that allowed them to use a variety of materials including everyday objects such as paper, cardboard, mirrors, and a web-based application to make their own data visualizations; the aim was to develop the children's DVL through construction of data visualizations (Bae et al., 2023). By qualitative analysis and evaluation, Bae et al. (2023) indicated that most of the children were able to create, analyze and draw meaning from their visualizations. The study by Angra and Gardner (2017) compared undergraduate students, graduate students and professors in light of their processes and performances when they constructed graphs on paper. Expert–novice differences were identified, such as more extensive planning, data transformation, and graph choice based on questions or hypotheses by experts, compared to minimal to no planning, raw data use, and intuitive reasoning by novices who were undergraduate students with little research experience (Angra & Gardner, 2017). However, both studies focused on constructing graphs or visualizations on paper or using physical materials. New challenges may arise when using computer-based data visualization tools. Moreover, these studies on visualization construction placed little emphasis on explicit strategies for construction of data visualizations. For example, Angra and Gardner's (2017) study did not include strategy use, but suggested incorporating strategic scaffolding

into future studies for the data visualization learning process. This current study addressed this issue by systematically identifying strategies for data visualization.

Theoretical perspectives for investigating data visualization

The cognitive model of drawing construction (CMDC) specifies the cognitive process when learners are asked to make drawings or visual representations for the purpose of learning, which may be applied to the case of data visualization. The CMDC suggests the importance of prior knowledge and cognitive processes including selecting, organizing and integrating during the construction of drawings or visualizations (Van Meter & Firetto, 2013). A three-phase self-regulation cycle has also been proposed to denote the role of metacognition during the construction process. For example, the task of drawing may trigger students' awareness that content is not well understood or visualization is not good enough; therefore, metacognitive self-regulation is needed to address the identified weakness in the process of drawing (Van Meter & Firetto, 2013).

Despite the process of data visualization mainly referring to the cognitive activities involved in the action of visualization, the perspectives of metavisualization (Gilbert, 2005, 2008, 2010; Justi et al., 2009) and metarepresentational competence (diSessa, 2004) underscore the importance of metacognition as a key to successful visualization. It is suggested that "metacognition in respect of visualization" be referred to as "metavisualization" (Gilbert, 2005, p. 15) and "a fluent performance in visualization" requires "metavisualization" (Gilbert, 2008, p. 5). To achieve metavisualization, students need to attain the stage of monitoring and controlling the cognition and learning associated with the visualization process (Gilbert, 2005). For example, learners need to become aware of monitoring their learning from visualizations, and execute control during the process of visualization such as retrieving, retaining and revising related images (Gilbert, 2005). When it comes to constructing data visualizations, metacognition involves regulating the process, such as planning the action, monitoring the progress, and evaluating the product of data visualization. Based on the theoretical perspective, this is crucial for facilitating a successful construction of data visualizations.

Similarly, the perspectives of metarepresentational competence emphasize metacognition and reflection so that the problematic aspects of representation, such as creating and choosing representations, and judging the value and adequacy of certain representations, can be dealt with (diSessa, 2004). An instance of metarepresentational competence in data visualization is an individual's capability to reflect on what constitutes a good

data visualization and to apply this insight in identifying flaws and limitations in the data visualization, aiming for improvement. In summary, these theoretical perspectives all indicate the importance of metacognition during the process of visualization. However, little empirical research has investigated how and what kinds of metacognition may be associated with data visualization. Investigating this issue would lead to extension and elaboration of the cognitive model with clearer delineation of the metacognitive types and processes, and of the metavisualization perspectives in the case of data visualization, which reveals critical mechanisms for successful data visualization, and provides insights for STEM education in terms of how to facilitate students' data visualization literacy.

Metacognitive and metavisual strategies

Metacognitive strategies can be defined as methods that one uses to facilitate and regulate one's own cognition to achieve certain goals such as task completion, which involves monitoring and controlling one's own learning (de Boer et al., 2018). Various types of metacognitive strategies have been identified by research in different subject areas of education. A review study identified three main kinds of metacognitive strategies in educational research, namely, strategies for planning and prediction, for monitoring and control, and for evaluation and reflection (de Boer et al., 2018). The three types correspond to the three main metacognitive regulation categories commonly indicated in the metacognition literature (Schraw & Moshman, 1995).

Ku and Ho (2010) examined 10 undergraduate students' uses of metacognitive strategies in tasks of hypothesis testing, verbal reasoning, argument analysis, understanding likelihood, and decision-making. They identified three types of metacognitive strategies, namely planning, monitoring and evaluating strategies, which are similar to the types identified in the review study of de Boer et al. (2018), and correspond to the metacognitive regulation categories as well. Ku and Ho also identified several sub-categories for each of the main categories, such as "inquiring task nature" and "inquiring task procedure" under the planning strategy.

Another study interviewed 26 undergraduate and 12 graduate students, engaging them in problem-solving tasks in organic chemistry to investigate their uses of metacognitive strategies during the interview (Blackford et al., 2023). The study identified 20 strategies, such as "set goals", "sort relevant info", "jot down ideas", and so forth, and categorized them into three main types, that is, planning, monitoring and evaluation strategies.

The research is clear about the three main types of metacognitive strategies that seem to be able to be

applied to all areas of learning and teaching. However, for effective use and concrete examples, identification of more metacognitive strategies at finer-grained levels for learning in various subject areas would help. For example, it is necessary to further reveal what metacognitive strategies there are at finer-grained levels, and how they are used when students are planning, monitoring and evaluating. This argument is consistent with the competence-based knowledge space theory which underscores the importance of considering a set of fine-grained skills specific to solving problems in a domain so that adaptive scaffolding in the learning environment can be developed at a later date (Heller et al., 2006; Steiner & Albert, 2011). It is also practically important for data science teachers to know and use specific and useful metacognitive strategies for data visualization construction in their teaching.

Specific to the learning of visualization, the use of metacognitive strategies has been referred to as the use of metavisual strategies (Locatelli & Arroio, 2014, 2016; Locatelli & Davidowitz, 2021), which involves a systematic series of planned or monitored actions for achieving a particular goal of visualization (Chang, 2022; Hung et al., 2021). Research has started to recognize the importance of teaching students to use metavisual strategies (Locatelli & Arroio, 2014, 2016; Locatelli & Davidowitz, 2021), and to identify several metavisual strategies that may lead to better visualization products in the case of visualizing science concepts and models (Chang, 2022; Hung et al., 2021; Locatelli & Arroio, 2014, 2016; Locatelli & Davidowitz, 2021).

This current study applied the coding framework of metavisual strategies by Chang (2022) and Hung et al. (2021) to examine how well the participants of this study demonstrated metavisual strategies during the data visualization task. Moreover, this study further identified and systematically analyzed non-metavisual strategies such as the personal preference strategy and student difficulties specific to the context of data visualization. ENA (Marquart et al., 2018; Shaffer et al., 2016; Wooldridge et al., 2018) was employed to generate students' network models showing the co-occurrences of which difficulties were accompanied by which strategies, to better understand patterns demonstrated during the process of students' data visualization construction.

Methods

The methodology that guides the methods and analyses of this study is quantitative ethnography (Kaliisa et al., 2021; Ruis & Lee, 2021; Shaffer, 2017). Ethnography focuses on investigating "process" to explain why and how individuals do "things that make up the range of human experience" (Shaffer, 2017, p. 31). Quantitative ethnography combines qualitative and quantitative methods in that

it suggests statistical techniques to analyze qualitative codes from big data as well as small data (Kaliisa et al., 2021; Ruis & Lee, 2021). Specifically, epistemic network analysis (ENA) was employed in this study, as detailed in the “Data coding and analysis” section.

Participants

Thirty first-year college students (26 females) and 27 10th-grade students (23 females) volunteered to participate in this study. They were recruited from a public university and a high school in Taiwan. Each of the participants received remuneration of NTD\$100 (about USD\$3.33) after participation. The participant selection criterion was indicated in the recruitment statement; that is, we only recruited first-year college and 10th-grade students, given that first-year college students have completed the 12-year basic education required in Taiwan, and 10th-grade students have completed their education at the junior high school level (Grades 7–9). The study and its procedure were approved by the Research Ethics Committee at National Taiwan Normal University (approval no. 201905HM040).

Eleven of the college students majored in science or had taken advanced science courses at the high school level. The 27 10th-grade students had not yet chosen a science or non-science major, but five of them had just started to take some advanced science courses at the senior high school level. Nevertheless, all of them had completed fundamental science courses at the junior high school level as is required by our national curriculum standards. The participants used computers frequently, mostly desktop and laptop computers, mobile devices and smart cellphones for 2–4 h daily. However, none of them had used any data visualization tools before (including the one used in this study, Tableau Public) or had taken any courses on data visualization.

Procedure

Each of the participants was individually engaged in the instruction, practice task, main task and retrospective interview. Detailed instruction and interview protocols were developed. In summary, the participants were instructed and guided to learn to use the computer-based visualization tool during the instruction, and practiced creating data visualizations and conducting think-aloud during the practice task. They were not disturbed or guided in the main task so as to maintain the authenticity and validity of their think-aloud processes. They were interviewed after the main task by the interviewer following the protocols.

A data visualization tool on desktop computers, Tableau Public, was used to present the dataset for the participants to construct their data visualizations. Tableau

Public provides general functions of computer-based data visualization tools such as visual template selection, variable selection for designated axes or dimensions, and visualization formatting functions. Internet access was also provided in case the participants needed to search for information. Thinking aloud was required during the data visualization process. Instruction on how to use Tableau Public and perform think-aloud was given by the interviewer before the main task. The participant was given time to learn and practice creating data visualizations using the visualization tool with another dataset (data about student enrollment) and thinking aloud before the main task.

During the main data visualization task, a dataset consisting of air quality variables and values collected at six monitoring stations (sites) in Taiwan across three years was provided to each of the participants in individual interview sessions. The choice of air quality is because it is a topic that is highly relevant to many people's daily life in Taiwan. The air quality data used in this study were downloaded from the “Government Open Data” website (<https://data.gov.tw/>), a platform offering government-released data for public scrutiny and exchanges on various topics. We did not alter the data to maintain authenticity.

Nevertheless, among the available datasets on the “Government Open Data” website in the topic of air quality, the selection of the dataset used in this study was decided in a panel meeting with a high school science teacher, a college instructor specializing in data visualization, and a science education researcher. The panel assessed the dataset in light of the complexity level and agreed that the dataset was neither overly simple nor overly challenging for both high school and college students. The dataset comprises a total of 61 variables. These variables include sites, dates, longitude, latitude, PM2.5, and other chemical or ionized elements related to air quality. The dataset consists of 915 rows, indicating the values of air quality variables for sites on different dates over a span of 3 years.

The participants' task was to transform the dataset to data visualizations so that the participant would use the data visualization to suggest funding allocation for air quality improvement in Taiwan, hypothetically to the Environmental Protection Administration. The task and the whole procedure remained the same for both college and high school students, facilitating valid performance comparisons based on identical assignments and procedures.

After the main visualization task, a retrospective interview was conducted to ask the participant to reflect on the difficulties and strategies, with interview questions such as “What was hard for you when performing the

task?” “What knowledge, skills, or strategies did you use to help you complete the task?” The interview was audio recorded. All sessions were carried out without a time limit. The time for learning and practicing using the tool lasted about one hour, depending on the needs of the participants. Then it took about one hour on average for the participants to complete the main task and retrospective interview, resulting in about two to three hours for the whole process including instruction and practice. All stages of the process were video- and audio-taped, screen-captured and transcribed.

Data coding and analysis

Coding scheme for strategy

To identify the strategies the participants used during their data visualization process during the main task, a coding framework was generated based on Chang (2022) and Hung et al. (2021) as well as on the data in this study. The coding scheme was discussed, and agreement was reached by two science education researchers and one data visualization researcher in multiple meetings to establish its content and expert validity. In summary, seven types of strategies that the participants used during their data visualization process were identified and defined (listed in Table 1; example excerpts for the strategies are provided in Additional file 1), five of which were categorized as metavisual strategies since metacognition is performed with participants' attention to the purpose of the data visualization task and the difficulties encountered. The other two were primary but non-metavisual strategies since little metacognition was observed while the participant was employing these two strategies. All the transcripts and screen-captured videos or audio

recordings from the two sources of data, the think-aloud visualization task, and retrospective interview, were inspected to make sure that the strategy coding scheme included all strategies demonstrated or indicated by the participants.

Coding scheme for difficulty

A coding scheme for difficulty demonstrated during the main task (Table 2; example excerpts for the difficulties are provided in Additional file 1) was generated based on the data of this study. The procedure to establish the content and expert validity of the coding scheme for difficulty was the same as the procedure for the coding of the strategy. As with the checking and triangulation for coding strategies, all the transcripts and screen-captured videos or audio recordings from the two sources of data, the think-aloud visualization task, and the retrospective interview, were inspected to make sure that the coding scheme for difficulty included all difficulties demonstrated or indicated by the participants. A total of six types of difficulties were identified and defined (listed in Table 2).

Unit of coding and inter-coder reliability

The unit of the qualitative coding for strategy and difficulty is by each participant's episode. An episode has a beginning and an end, and consists of a series of sentences or actions that occur based on an identical reason or purpose and can be unified within one episode and distinguished from other episodes (van Dijk, 1981). Since the study focuses on students' strategies and difficulties that usually require a series of sentences or actions, it is therefore suitable to code the data based on the analysis

Table 1 Coding framework for strategy use during construction of data visualizations (adapted from Chang, 2022)

| Strategy | Definition |
|--------------------------------|--|
| Metavisual strategy | Systematic series of planned or monitored actions for achieving a particular goal of visualization |
| Focusing strategy | Performance indicated that the participant monitored progress by continuously matching among her/his current ideas, the expressed data visualization, and the goal of the task |
| Inducting strategy | Performance indicated that the participant used reflection and self-questioning to generate criteria that guided identification of important variable(s) needed in the data visualization task |
| Perfecting strategy | Performance indicated that the participant identified flaws or limitations in the initial data visualization and continuously thought about how to improve the quality of the data visualization to fulfill the task request and achieve the task goal |
| Resourcing strategy | Performance indicated that the participant retrieved existing conceptions or searched for online information to identify and obtain resources needed based on the purpose of the task |
| Deducing strategy | Performance indicated that the participant applied her/his own ideas or knowledge to guide the action for the data visualization task |
| Non-metavisual strategy | Series of actions with little or no evidence of metacognition |
| Personal preference | Performance indicated that the participant constructed a certain data visualization because she/he was familiar with that type of data visualization |
| Trial and error | Performance indicated that the participant tried several types of data visualizations and decided on a data visualization without being able to give a reason or consider the purpose |

Table 2 Coding framework for difficulties during construction of data visualizations

| Difficulty | Definition |
|--------------------------------------|--|
| Initial representation difficulty | The participant showed difficulty representing her/his ideas using a particular type of data visualization; the generated data visualization was different than what the participant wanted to express |
| Multiple representation difficulty | The participant was able to generate one data visualization, but was not satisfied. The participant tried to construct another data visualization complementary to the first data visualization generated but failed |
| Data simplification difficulty | The participant indicated a difficulty that there were too many data and it was difficult to simplify the data into a data visualization |
| Information comprehension difficulty | The participant indicated a difficulty comprehending the information presented in the question or in the data table |
| Technological difficulty | The participant showed difficulty using a feature or function of the data visualization technology, including using general functions of computer-based data visualization tools or features that afford newer representations than those on paper or using physical materials |
| Content knowledge difficulty | The participant indicated a lack of sufficient content knowledge on the topic of air quality to make a good data visualization |

unit at the episode level. For example, there are usually multiple episodes for a participant to complete the data visualization task, such as searching for the needed information on the Internet before starting to construct a data visualization (one episode), reading and comprehending the data table provided to determine the critical variables (another episode), and starting and completing the data visualization (the other episode).

The software, NVivo, was used to aid the coding and analysis processes. The episodes showing the occurrence of any of the strategies and difficulties in the coding schemes (Tables 1, 2) were highlighted and assigned the correspondent codes in NVivo. Note that the coding of the strategies and difficulties included multiple data sources, not only the think-aloud data but also the retrospective interview data and the process video of the participants' processes and products of constructing data visualizations. Two coders coded all 57 participants' data. The inter-coder reliability of the coding was 0.84 (Cohen's kappa). Inconsistent codes between the coders were discussed and resolved. Assertions were made by inspecting the coded data and results, and searching for confirming and disconfirming evidence.

Employment of ENA

ENA uses statistical and data visualization techniques to generate network models of nodes and connections that reveal how often the nodes occur, how the nodes are connected (i.e., co-occur), and how the occurrence and connection of the nodes might differ between groups of individuals (Shaffer, 2017). The nodes in this study consisted of the seven strategies and six difficulties listed in Tables 1 and 2, respectively. Our inspection of the coded data indicated that the participants often used more than one strategy accompanied with their difficulty during the process of data visualization. It is therefore suitable

and beneficial to employ ENA so that how the strategies and difficulties co-occurred could be investigated and revealed.

The ENA1.7.0 Web Tool was used to quantitatively process the codes in Excel format (Marquart et al., 2018). A codebook in Excel consisting of a list of the codes for all the participants was generated, with the occurrence of each strategy or difficulty coded as 1, and absence as 0. The data format for ENA is stanza-based. A stanza is defined as the recent temporal context associated with each line of data (Shaffer, 2014). For a specific line in the data, the stanza comprises other lines in the data that form part of the recent temporal context for that line (Shaffer, 2018). The ENA software implements "stanzas" by utilizing a moving window of a fixed size " w "; essentially, each line of data is connected with the " $w-1$ " lines of data in the conversation preceding the particular line, creating a total window size of " w " lines (Shaffer, 2018, p. 523). In this study, since there was no interaction among the participants, the stanza size w for the analysis was set to 1 (Shaffer, 2014; Zörgő et al., 2021).

ENA constructs one model for each participant. For example, Fig. 1 shows the structure of the network model of the participant ID#DVSH010. It shows that this high school student encountered the multiple representation difficulty, and employed resourcing and deducing strategies during the process of data visualization. Then, each network model can be represented as a single point in the figure, so that all of the participants' models can be represented and analyzed collectively in the model (Fig. 2). In Fig. 2, each point is the centroid of the corresponding network for each participant. Figure 2 shows all participants' centroids in the form of dots. The centroid is "the arithmetic mean of the edge weights of the network model" and is "much like the center of mass of an object" (Shaffer et al., 2016,

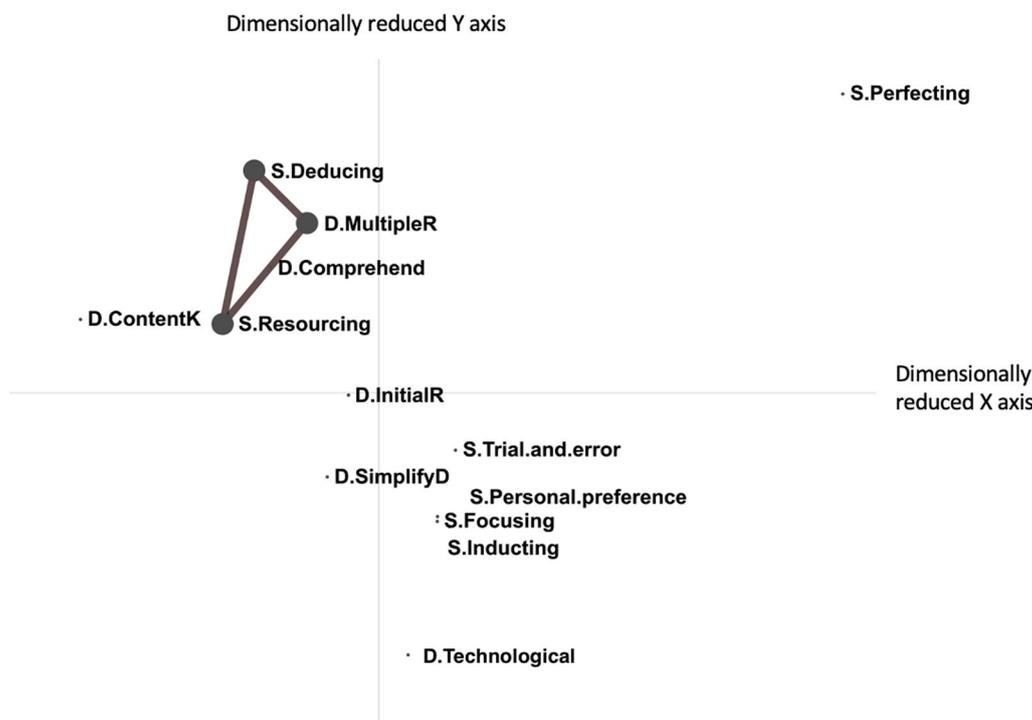


Fig. 1 The network model of DVSH010. S.Focusing: focusing strategy; S.Inducting: inducting strategy; S.Perfecting: perfecting strategy; S.Resourcing: resourcing strategy; S.Deducing: deducing strategy; S.Personal.preference: personal preference strategy; S.Trial.and.error: trial-and-error strategy; D.InitialR: initial representation difficulty; D.MultipleR: multiple representation difficulty; D.SimplifyD: data simplification difficulty; D.Comprehend: information comprehension difficulty; D.Technological: technological difficulty; D.ContentK: content knowledge difficulty

p. 16). The mean position of the points with the confidence interval can then be calculated to indicate the whole group’s performance (Shaffer et al., 2016). The black square in Fig. 2 is the mean position, and the box around the square indicates the 95% confidence interval. Meanwhile, the network model demonstrated by all the participants collectively is represented in Fig. 3. A larger sized node indicates that more participants demonstrated this strategy or difficulty. A thicker line indicates that the connection between the two nodes occurred more frequently.

Moreover, ENA allows comparison between groups, such as the high school and college students in this study. The mean positions with the confidence intervals for the two groups were calculated, and an independent *t* test was performed to indicate whether the two groups’ models were significantly different in terms of the positions on the X and Y dimensions. Subtracted network models were also employed in this study, which subtracted the weight of each connection in one group’s model from the corresponding weighted connection in the other group’s model, to visualize how the structure of the two groups’ models differed (Shaffer et al., 2016). To better visualize the differences, the scale for edge weights was set to 2 in the ENA1.7.0 Web Tool.

Results

Overview of data visualization strategies and difficulties (RQ1)

Overall, about 91% (52 of 57 students) constructed at least one adequate data visualization, although to varying degrees of proficiency and accuracy. The evaluation of the data visualization products was based on detailed scoring rubrics with procedures to ensure the validity and reliability of the evaluation. In general, the scoring rubrics focused on the suitability of the chosen visualization type for the selected data values and the presence of sufficient and accurate information as forms of representations to accomplish the purpose of the visualization task. As previously mentioned, the majority of the students constructed satisfactory data visualization products, but delving into further details on the scoring goes beyond the scope of this paper.

Despite the satisfactory visualization products generated by the majority of the students, during the process of data visualization, they still encountered difficulties and used strategies to overcome them. Table 3 provides an overview of the difficulties and strategies demonstrated by the participants. During the data visualization task, the most common difficulty, as demonstrated by 21 of the 57 participants, was the technological difficulty. The

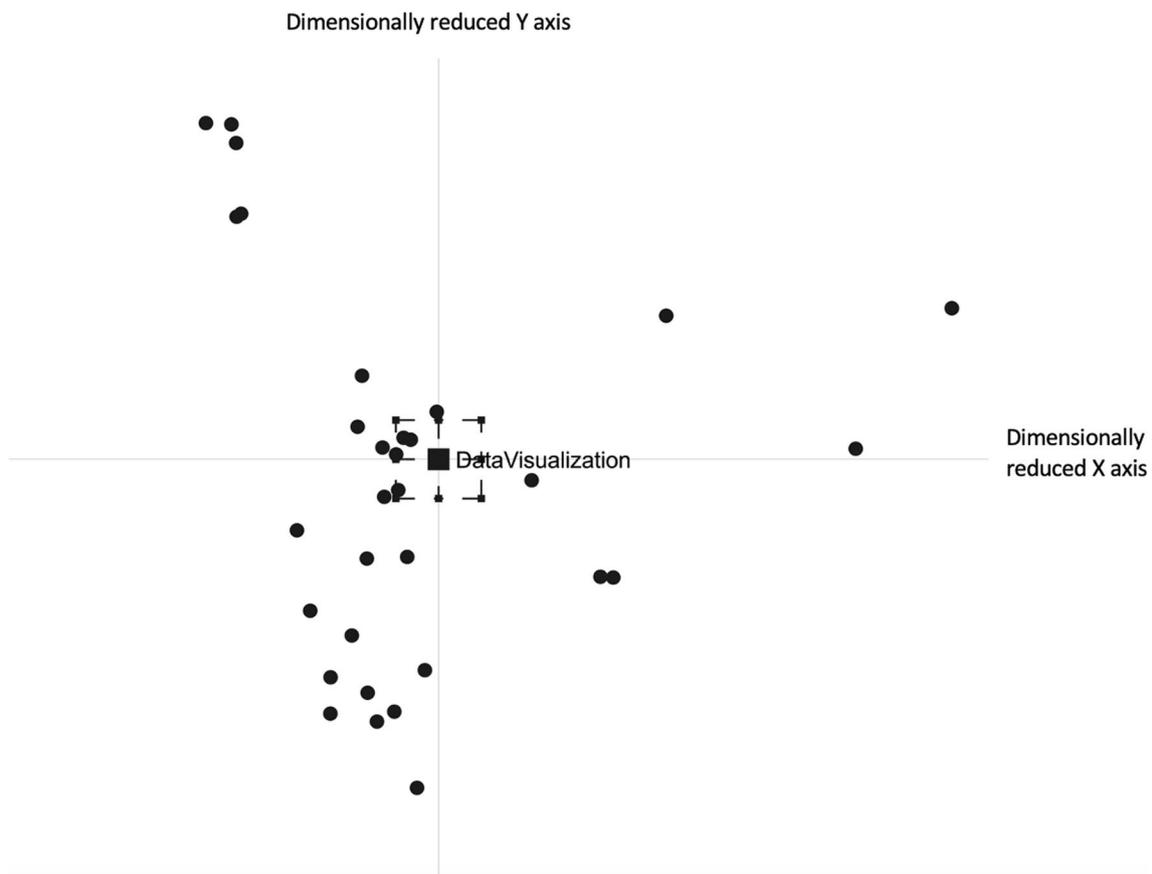


Fig. 2 All participants' centroids (dots) with the mean (square) and 95% confidence interval (box) of all participants' centroids

technological difficulty arose from constructing a data visualization using a computer-based data visualization tool as opposed to on paper or using physical materials such as cardboard. Computer-based data visualization tools provide new affordances not available using the paper-based or physical object methods, but such new affordances may also introduce difficulties. The students showed technological difficulty in terms of using visual templates (e.g., wanting to create a 3-D visualization but failing [ID#DVCS011]), assigning variables (e.g., including three variables for a visual template, but the resulting visualization was useless due to inadequate variable assignment [ID#DVCS003]), or formatting visualizations (e.g., wanting to show and hide the latitude and longitude lines on a map representation [ID#DVCS001]).

The second most common difficulty relates to content knowledge difficulty. The participants expressed that they were unsure about which elements in the air were most detrimental to humans' health. Also, 17 of the participants demonstrated the initial representation difficulty, indicating a lack of representation skills, and 11 indicated the difficulty of simplifying the data for visualization.

Only three participants indicated the difficulty of comprehending the information presented in the task or in the data table.

Moreover, Fig. 3 presents the ENA result showing the mean network model of the participants in light of how the difficulties and strategies co-occurred with each other. It is evident that the difficulties and strategies did not occur in isolation. Rather, a central pattern revealed in Fig. 3 is that the inducting and resourcing strategies were used by the students to overcome the technological and content difficulties. Moreover, it is revealed in the figure that the technological difficulty may co-occur with the content or other types of difficulties.

In terms of strategy use, 27 of the 57 students used the inducting strategy. The second most frequently used strategy was the resourcing strategy, followed by the deducing strategy. Ten of the participants used the perfecting strategy, and four used the focusing strategy. Moreover, eight participants used the trial-and-error strategy and five used the personal preference strategy, both of which are non-metavisual strategies. The meanings of these strategies and how they were applied in the

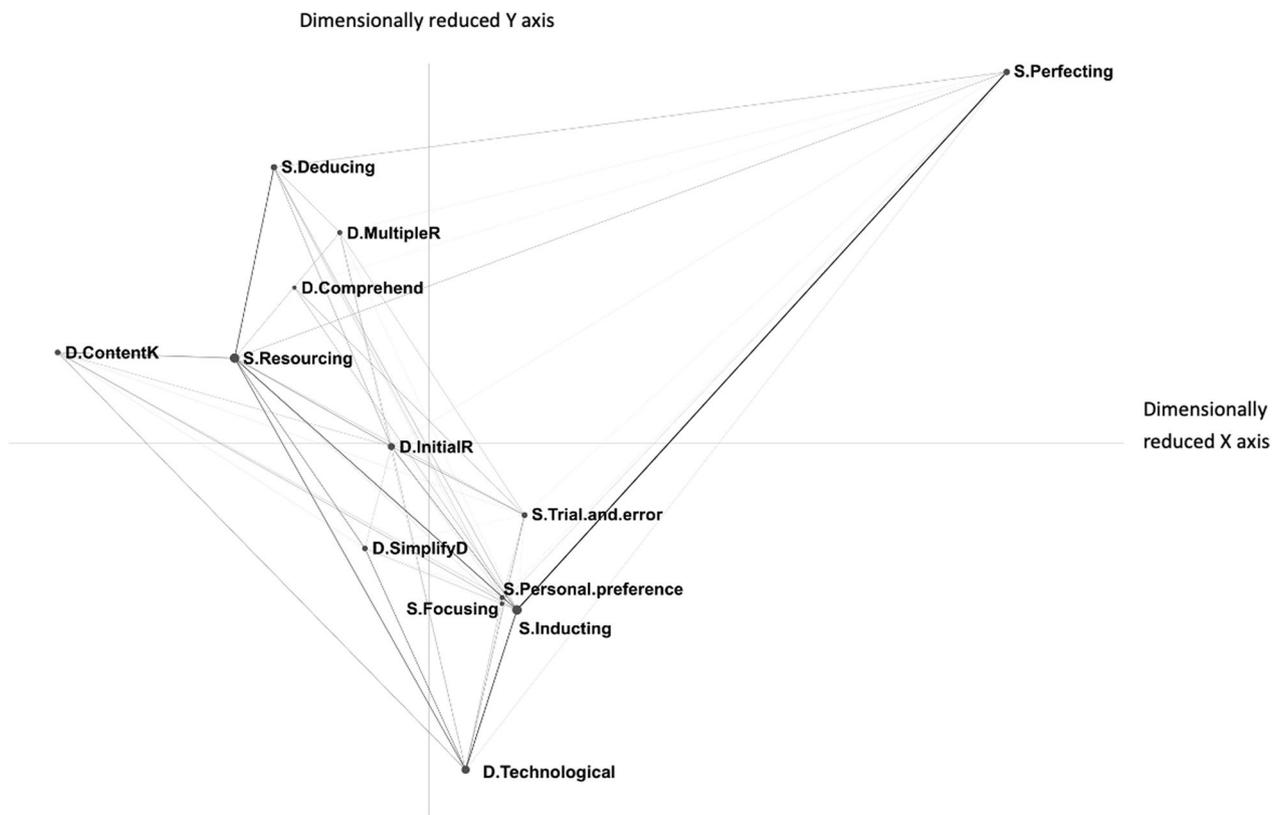


Fig. 3 The mean network model demonstrated by all of the participants. S.Focusing: focusing strategy; S.Inducting: inducting strategy; S.Perfecting: perfecting strategy; S.Resourcing: resourcing strategy; S.Deducing: deducing strategy; S.Personal.preference: personal preference strategy; S.Trial.and.error: trial-and-error strategy; D.InitialR: initial representation difficulty; D.MultipleR: multiple representation difficulty; D.SimplifyD: data simplification difficulty; D.Comprehend: information comprehension difficulty; D.Technological: technological difficulty; D.ContentK: content knowledge difficulty

Table 3 Overview of the difficulties and strategies demonstrated during the process of data visualization

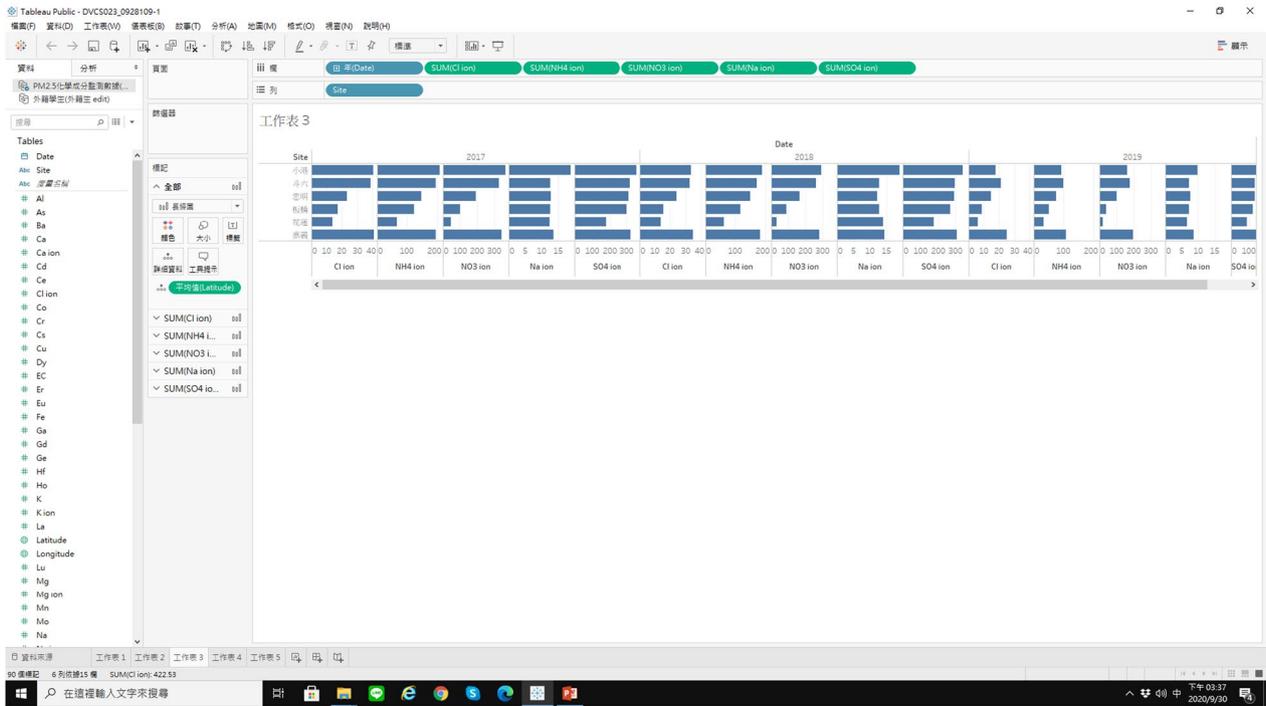
| Performance | n (% = n/57) |
|--------------------------------------|--------------|
| Difficulty | |
| Initial representation difficulty | 17 (29.8%) |
| Multiple representation difficulty | 6 (10.5%) |
| Data simplification difficulty | 11 (19.3%) |
| Information comprehension difficulty | 3 (5.3%) |
| Technological difficulty | 21 (36.8%) |
| Content knowledge difficulty | 18 (31.6%) |
| Strategy | |
| Focusing strategy | 4 (7.0%) |
| Inducting strategy | 27 (47.4%) |
| Perfecting strategy | 10 (17.5%) |
| Resourcing strategy | 21 (36.8%) |
| Deducing strategy | 16 (28.1%) |
| Personal preference strategy | 5 (8.8%) |
| Trial-and-error strategy | 8 (14.0%) |

context of data visualization are delineated in the following section.

High-level-, basic-level-, and non-metavisual strategies for data visualization (RQ2)

The use of the focusing, perfecting, and inducting strategies demonstrates a high level of metavisual strategy since metacognition is constantly employed. For example, the participant, ID#DVSH044, constantly asked himself “What is the purpose of the task again?” during the data visualization process to remind himself to stay focused on achieving the goal of the task and to monitor his progress. He was therefore coded as using the focusing strategy. The perfecting strategy also requires monitoring and reflecting on the quality of the data visualization. By identifying the flaws or limitations of the current data visualization, some participants continued to improve their visualizations by constructing another data visualization that would be complementary to the original one, and hence were coded as using the perfecting strategy. For example, the participant, ID#DVCS023, indicated

(a)



(b)

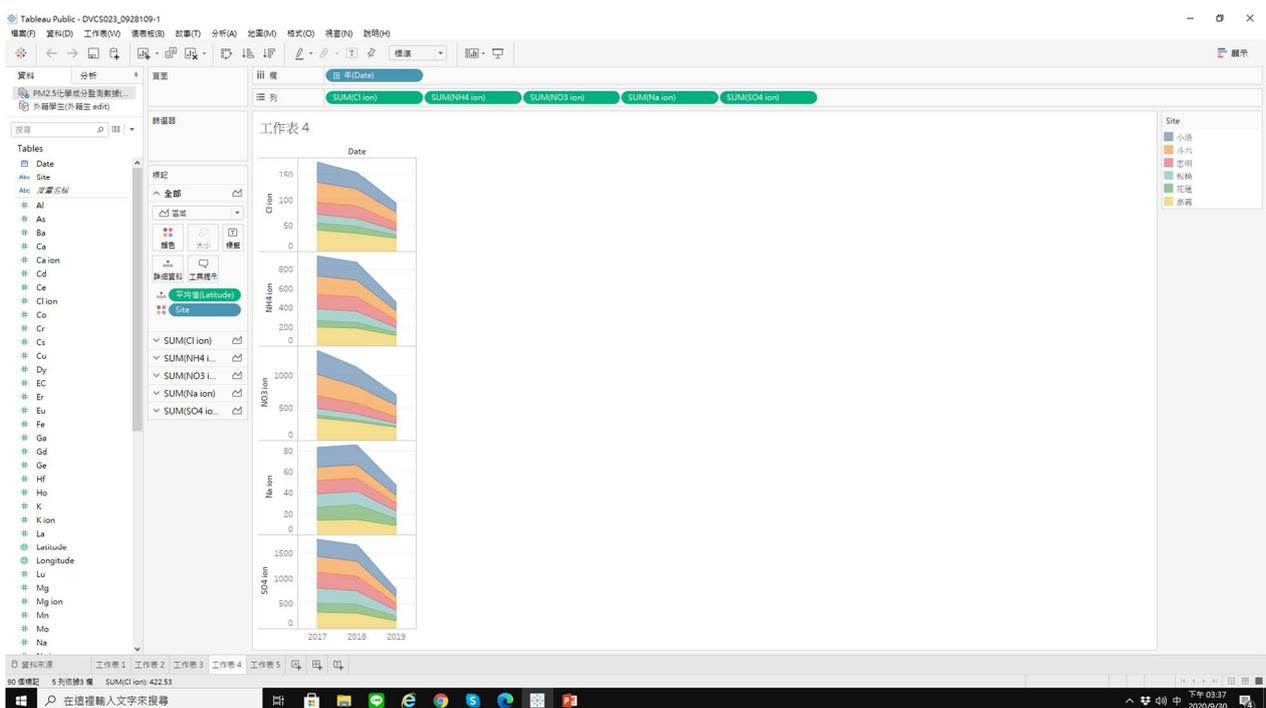


Fig. 4 Example student-generated data visualizations (ID#DVCS023). 4(a): upper, the first data visualization created by DVCS023; 4(b): lower, the second data visualization created by DVCS023

that her data visualization did not clearly show the relationships between two variables (Fig. 4a), so she wanted to “make a new graph and then see if it can clearly show the relationship of the data.” She then successfully made a second data visualization (Fig. 4b) to complement her first one.

The inducting strategy also requires substantial meta-cognition, and is a very useful strategy demonstrated by the participants to make progress in the process of data visualization, when there is a large amount of data and the participant at first has not had clear or formed conceptual frameworks to guide the data visualization process. The following excerpt from ID#DVCS026 shows an example of using the inducting strategy.

DVCS026: If the data are for the government to make decisions about which location should receive more funding, I feel that it seems meaningless to compare among the dates. I should focus on the location...such as those [the values] that are so small across the locations, I will first remove them because it seems meaningless to compare these small numbers across the sites...now I need to think about how to present the changes by years across the sites...I think I will use this presentation, first the sites then the years to see the changes...But I think right now the content is still too much, and I want to present

the information in one chart, so I am going to remove more variables that have very small numbers.

Through self-questioning and reflecting, the participant ID#DVCS026 was able to generate the criteria that guided her construction of the data visualization. She decided to focus on the variables that addressed the purpose of the task, and variables that were critical to indicate the air quality across the sites. She removed those variables based on her criteria that they were not critical (all with small numbers across the sites) and were not relevant to the purpose. These criteria were not formed at first before the task; rather, they were formed as a result of DVCS026’s inspection and observation of the data (Fig. 5 shows the data visualization created by DVCS026). This bottom-up approach is therefore called the inducting strategy. The constant self-questioning and reflection indicate employment of metacognition.

The use of the resourcing and deducing strategies demonstrates the basic level of metavisual strategy since metacognition is needed, but is implicitly and less constantly employed. For example, the participant ID#DVSH010 checked and confirmed the purpose of the task and searched online for information based on the purpose: “I think the government will allocate the funding based on the degree of air pollution, so I am going to search for information regarding degrees of air pollution

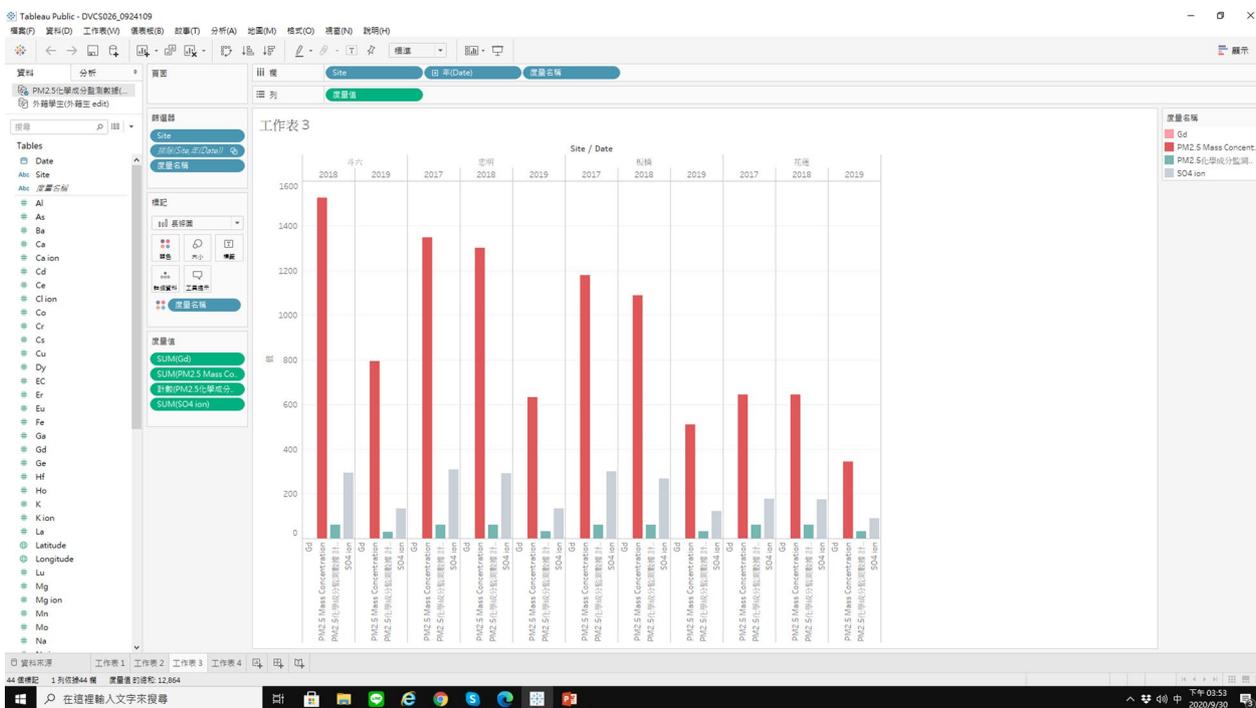


Fig. 5 An example of a student-generated data visualization (ID#DVCS026)

and which elements and the amount of them will cause different degrees of air pollution" (ID#DVSH010). The search behavior of DVSH010 was guided by her keeping in mind the purpose of the task; hence, metacognition was employed; this strategy is called the resourcing strategy. Another strategy, the deducing strategy, also involves the participant keeping in mind the purpose of the task, and the data visualization is fluently conducted and completed following the participant's own ideas or knowledge with less reflecting or self-questioning, as the following excerpt shows, which indicates that the participant completed the task in a top-down manner in that she applied her knowledge to form the data visualization that she wanted, which is therefore called the deducing strategy.

DVCS009: I want to know at every location by each year, how much the element, Cr, was detected. I put year and location as the rows, and the element Cr as the column. Then I use a bar chart. Now I see that in every year the amount of Cr detected is always the highest in Xiaogang.

In summary, 43 out of the 57 participants (75%) demonstrated the basic or high-level metavisual strategies. This result indicates that the task of constructing data visualizations may prompt the majority of the students to engage in the use and practice of metacognition. Nevertheless, there were four students who used only the non-metavisual strategies, and 10 students did not demonstrate any of the strategies; this finding is discussed as follows.

The personal preference and trial-and-error strategies are non-metavisual in that they involve very little metacognition. The participant, ID#DVCS022 showed an example of using the trial-and-error strategy. During the process of constructing the data visualization, she first selected two variables, date and site, and defined one as one variable in the column and the other as one variable in the row. She then said, "Now I am thinking how to put all the chemicals (the other variables) into the table...I will just select several of them randomly, and then change the type of graph." In the retrospective interview when she was asked about the strategies she used, she indicated, "I think the best way is just to try everything out. I tried every type of graph and diagram to find the one that looks good to me." The data from both the think-aloud and retrospective interview triangulated that DVCS022 was using a trial-and-error strategy. Although she indicated that she wanted to "try everything out", she in fact only tried out several types that she was familiar with. Therefore, DVCS022 also demonstrated a personal preference strategy, indicating "the bar graph is the type that I am most familiar with", when she was asked why she used the bar graph. The criteria for using and

determining the data visualization are mainly based on trials and personal criteria such as "looking good to me" or being "most familiar", as opposed to considering the purpose of the data visualization for the task.

Ten students were not coded as demonstrating any of the strategies. Eight of them showed a tendency to avoid effort by finishing the task quickly, with little evidence of thinking or reasoning during the process of data visualization. The observation of their actions and evaluation of their products (i.e., the data visualization) indicated that they very quickly constructed only one acceptable (but not the best) data visualization. In the retrospective interview, they were not able to articulate any strategies, but responded vaguely, for example stating "I wanted to show a clear chart" [ID #DVSC015]. The other two students encountered difficulties but were unable to come up with strategies on their own to overcome the difficulty, which prevented them from successfully completing the visualization task. Therefore, explicit support or scaffolding for using strategies is needed to address the needs of these students.

Comparisons of college and high school students' network models of strategy and difficulty during the process of data visualization (RQ3)

Figure 6 shows the distribution of the centroids by the college (red dot) and high school (blue dot) students. It should be noted that the dots are positioned in the reduced and rotated dimensions using the visualization technique in ENA (Shaffer, 2018), aiming to emphasize the significant differences between the college and high school students. Judging from the positions of the means and the 95% confidence intervals of the two groups in Fig. 6, the mean models of the two groups differed significantly. The t -test results further confirmed that the models produced by the college and high school students significantly differed on the X-dimension ($t=6.32$, $p<0.001$), but not on the Y-dimension ($t=0.00$, $p=1.00$). An overall pattern is that the college students' models were distributed towards the right-hand side of the dimensions, whereas the high school students were towards the left-hand side (Fig. 6).

The subtracted model (Fig. 7) shows how the models differed. According to the nodes, the college students' models distributed towards the right-hand side of the dimensions focused more on the technological and data simplification difficulties, whereas the high school students' models distributed towards the left-hand side focused more on the initial representation, multiple representation, comprehension and content knowledge difficulties.

Moreover, in Fig. 7, connections colored red were stronger in the college students' models, whereas

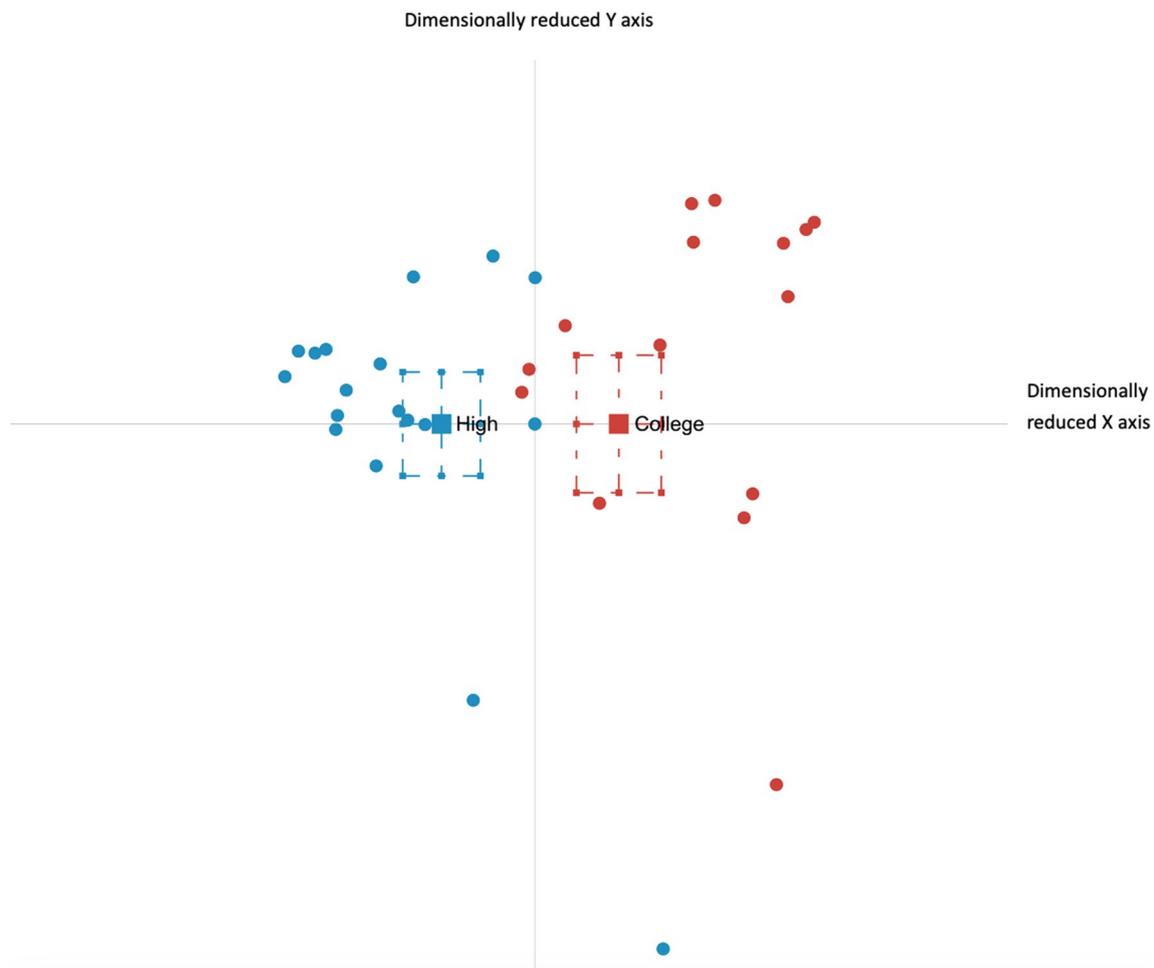


Fig. 6 (Color online) The college (red dot) and high school (blue dot) students' centroids. The red square and box represent the mean and 95% confidence interval for the mean of the college students' centroids, respectively, and the blue square and box represent the mean and 95% confidence interval for the mean of the high school students' centroids, respectively

connections colored blue were stronger in the high school students' models. Connections in this study mean co-occurrence of two connected nodes, which can indicate how encountered difficulties were resolved by which strategies, or how certain types of difficulties or strategies were related to each other. Figure 7 indicates that, compared to the high school students, the college students more often used the inducting, resourcing, and perfecting strategies to solve the problems caused by the technological difficulty and the difficulty to simplify the data for representation. It was observed that more college students focused on the challenge of how to simplify the data and to perfect the data visualization for best representations. The technological difficulties they encountered often related to their intention to use more and advanced functions to construct data visualizations that would best show their ideas. It is also revealed in Fig. 7 that, more for the college students, the technological

difficulty was related to the data simplification and sometimes to the content knowledge difficulty.

In comparison, the high school students used the resourcing and deducing along with the non-metavisual strategies, including the trial and error and personal preference strategies, to solve the problems they encountered. Specifically, for the initial representation difficulty, more high school students used the resourcing and trial-and-error strategy, and some used the personal preference strategy. For the content knowledge difficulty, which was often linked to the initial representation and information and comprehension difficulties, the students used the resourcing, focusing and trial-and-error strategies. For the multiple representation difficulty, which was sometimes related to the initial representation difficulty, the students used the resourcing, deducing and trial-and-error strategies. An overall trend revealed in the figures between the college and high school students was that more college

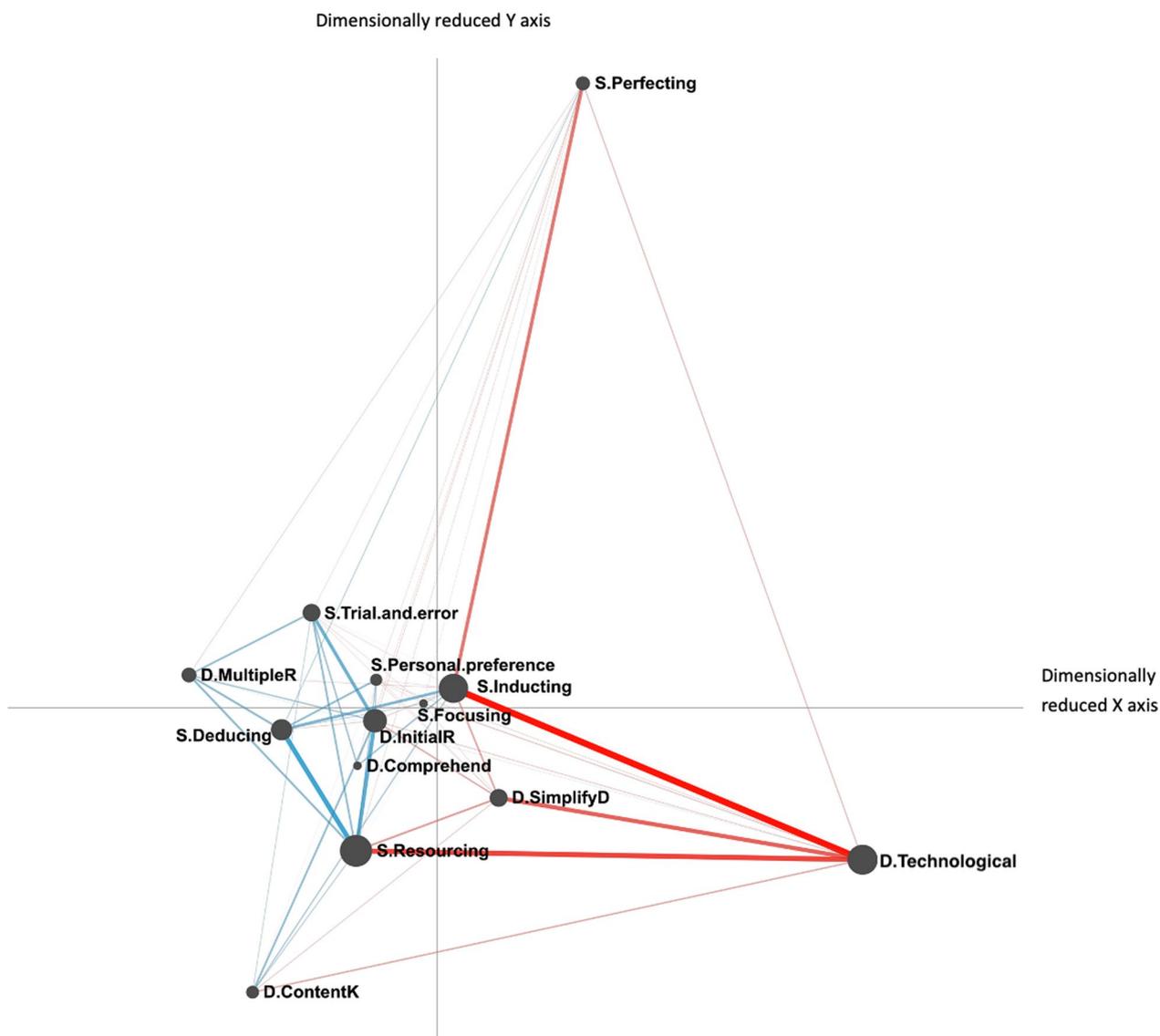


Fig. 7 (Color online) The subtracted network model between the college (in red) and high school (in blue) students. S.Focusing: focusing strategy; S.Inducting: inducting strategy; S.Perfecting: perfecting strategy; S.Resourcing: resourcing strategy; S.Deducing: deducing strategy; S.Personal.preference: personal preference strategy; S.Trial.and.error: trial-and-error strategy; D.InitialR: initial representation difficulty; D.MultipleR: multiple representation difficulty; D.SimplifyD: data simplification difficulty; D.Comprehend: information comprehension difficulty; D.Technological: technological difficulty; D.ContentK: content knowledge difficulty

students used the high-level metavisual strategies such as the inducting and perfecting strategies, whereas more high school students used the basic-level metavisual strategies such as the deducing strategy, and non-metavisual strategies such as the personal preference and trial-and-error strategies.

Discussion

Metacognition associated with construction of data visualization

Theoretical accounts of learning by visualization recognize the importance of metacognitive processes, but have done little to clarify what those processes might be

(Van Meter & Firetto, 2013). Similarly, the process of data visualization construction and interpretation suggested by Börner et al. (2019) has also focused on the cognitive or behavioral processes, including acquiring, analyzing, visualizing, deploying and interpreting. This current study provides evidence for the importance of metacognition by identifying metacognitive strategies for successful data visualization. These metacognitive strategies demonstrated by the majority of the participants indicate the important role of metacognition that is associated with and also needed during the process of constructing data visualizations. Specifically, it was identified in this study that the metacognitive process may include self-questioning, monitoring and reflecting. The metavisual strategies associated with data visualization may include inducting, focusing, perfecting, resourcing and deducing strategies.

Research has been clear about three main types of metacognitive strategies, namely planning, monitoring, and evaluating (Blackford et al., 2023; de Boer et al., 2018; Ku & Ho, 2010). This study builds on these findings to further identify metacognitive strategies at finer-grained levels, so that these strategies provide concrete examples of how to employ metacognition in the subject area of data visualization. Recent research has started to incorporate metavisual strategies in instruction to teach students using metavisual strategies for visualization tasks in science (Locatelli & Arroio, 2014, 2016; Locatelli & Davidowitz, 2021). Future research may consider choosing and incorporating various metavisual strategies identified in this study into instruction for the learning of data science and different subject areas, and for testing the effects.

Learning difficulties for constructing computer-based data visualizations

Research has identified three types of challenges faced by students constructing computer-based data visualizations, namely decomposing questions and goals into data attributes, designing visual mappings, and interpreting visualizations (Grammel et al., 2010). In comparison, the six types of difficulties identified in this study may provide insights into the reasons why challenges relating to the visualization construction may occur, or the factors contributing to the challenges. For example, the first challenge, decomposing questions and goals into data attributes, can stem from the information comprehension, content knowledge and data simplification difficulties. The second challenge, designing visual mappings, may be related to the initial representation, multiple representation and technological difficulties.

Moreover, research has found that designing visual mappings was the most problematic step demonstrated

by students, referring to the difficulty of selecting adequate visual templates (Grammel et al., 2010). Similarly, another study identified a common difficulty of constructing a data visualization, which is the inappropriate or poor choice of visualization types for particular datasets, causing the problem of the constructed visualization failing to convey clear meaning or not being aligned with the purpose (Chrysantina & Sæbø, 2019). In this current study, we found that the most demonstrated difficulty was the technological difficulty, including using the features and functions of the application tool to select visual templates, assign variables, and format visualizations. Moreover, based on the difficulties identified in this current study, we may further provide insights into the possible factors causing this common problem, including students' lack of knowledge needed to comprehend the information in the dataset and the task (i.e., the content aspect, which would be the knowledge of data and the topic of air quality in this study), and the computer-based visualization tool (i.e., the technology aspect), and a lack of representation skills to generate initial and multiple representations (i.e., the representation aspect).

Learning strategies for constructing computer-based data visualizations

Research indicates that novices often rely heavily on their prior experiences with data visualization types when constructing a data visualization (Grammel et al., 2010). This current study found five students who used a similar strategy, which is called the personal preference strategy in this study. They chose to construct a certain type of data visualization only because the data visualization was most familiar or looked good to them, without being able to further provide other reasons at the time of using the strategy.

Another non-metavisual strategy, the trial-and-error strategy was used by eight students in this study. Researchers argue that all programming is reasoned through the trial-and-error strategy if the definition of this strategy refers to a method to explore decision trees that involve considering all possible solutions (Merisio et al., 2021). However, in this study, we have defined the trial-and-error strategy as a strategy demonstrated by the participant who randomly, arbitrarily, and purposelessly tries out some data visualization types, as opposed to systematically trying out all possible solutions. Moreover, the use of ENA to generate the network model indicates a result that the trial-and-error strategy was more often used by the high school students who encountered the difficulty of generating an initial data visualization. The model also shows that participants may combine the use of the trial-and-error strategy with other strategies, such as the resourcing strategy.

Therefore, the argument we would like to make is that rather than treating the non-metavisual strategies as inappropriate or undesirable, we think that these strategies are intuitive and can be useful as building blocks for students to make progress and develop other strategies. However, students who used only these non-metavisual strategies showed limitations in the process of constructing data visualizations and using the visualizations for scientific reasoning. Moreover, there were students in the study who encountered difficulties but were unable to generate any strategies on their own. The metavisual strategies identified and exemplified in this study including the focusing, inducting, perfecting, resourcing and deducing strategies can be incorporated into instruction to teach students about these strategies.

Designing interventions to promote data visualization literacy

Research on teaching data visualization in formal education has proposed various teaching methods or interventions, such as the three-phase data visualization process (Camm et al., 2023) and the seven data visualization stages (Byrd & Dwenger, 2021). These methods primarily draw on theoretical and epistemological perspectives, advocating a top-down approach for designing interventions or instruction. In comparison, the current study employed an equally important bottom-up approach for instructional design. Specifically, it identified students' challenges and strategies relating to data visualization, offering insights into designing learning support and scaffolds tailored to students' needs.

Constructing data visualizations is an important aspect of DVL, but relatively few studies have addressed this aspect (Börner et al., 2019). By systematically investigating the learning difficulties faced by the 57 students, the study proposes three core aspects for consideration when designing teaching support and scaffolds to promote students' DVL, especially for data visualization construction, namely the content, representation and technology aspects. That is, students need support in developing sufficient content knowledge of data for a certain topic (such as data attributes for the topic of air quality), and representation skills to precisely represent data and ideas in visual formats. Also, with easy access to computer-based data visualization tools which provide new affordances, it is becoming equally important to support students' development of knowledge of the application tools.

Instrumentation theory emphasizes the mediating role of a tool in learning since the affordances and constraints of a tool may influence students' learning processes and strategies (Doorman et al., 2012). Therefore, supporting students in constructing data visualizations via a computer-based data visualization tool

may require substantially different interventions or instruction than those on paper or when using physical objects. Specifically, the newly identified technological difficulty in this study indicates the need for students to develop knowledge of the application tool for data visualization. This kind of knowledge is called technological content knowledge, and refers to the "knowledge about the manner in which technology and content are reciprocally related" (Graham, 2011, p. 1954). Interventions are needed to support students in integrating knowledge of topic data, representation, and technology affordance, as the cognitive model of drawing construction (CMDC) suggests that prior knowledge plays an important role in the construction process (Van Meter & Firetto, 2013).

Research has found that college and high school students' DVL may be significantly different in the aspect of comprehending and interpreting data visualizations (Binali et al., 2022). This is likely due to differences stemming from past experiences as college students may be more often exposed to the Internet, social media and diverse courses that provide opportunities for interacting with data visualizations (Angra & Gardner, 2017; Binali et al., 2022; Mansoor & Harrison, 2018). In exploring the differences in college and high school students' processes of constructing data visualizations, this study evidenced significant differences through ENA. More college students in this study focused on the difficulty of how to simplify the data with the intention to generate the best data visualization. This led to the need to use more and advanced functions of the computer-based data visualization tool. With the rapid development of computer technology, data visualizations can be easily generated using computer visualization tools (Li, 2020; Unwin, 2020). However, the college students still needed support and guidance to take full advantage of the functions and features of the visualization tool, as the results showed that the technological difficulty was the most frequent difficulty encountered. Future research needs to design instruction that addresses the different needs of college and high school students in order to promote students' DVL, and should test the effectiveness of the instruction.

Moreover, the study goes beyond identifying difficulties and comparing the differences between the college and high school students to also systematically investigating the demonstrated strategies. The identified metavisual strategies and how they connected to the difficulties via ENA provide concrete examples for incorporating learning strategies to facilitate students' development of DVL, especially for constructing data visualizations using application tools.

Innovative use of ENA to reveal how difficulties are resolved with strategies

ENA is an innovative and established method that can be used to analyze and model individuals' cognitive work or conceptual frameworks (e.g., Chang & Tsai, 2023; Rachmatullah & Wiebe, 2022), or groups' collaborative discourse, discussion, action or behavior (e.g., Bressler et al., 2019; Sun et al., 2022; Zhang et al., 2022). The use of ENA is based on the perspective that students' ideas or actions, no matter whether working individually or collaboratively, rarely occur in isolation (Linn et al., 2004). This current study applied ENA to model and visualize the co-occurrence of the difficulties and strategies so that how the participants demonstrated the difficulties and how they used strategies to resolve the difficulties during the practice of data visualization could be investigated. Such application of ENA enables results and insights that may not be realized when only codes are counted separately without consideration of their co-occurrence.

Conclusions

Developing students' data visualization literacy (DVL), such as the ability to adequately construct and interpret data visualizations through which conclusions are made and communicated, has increasingly become an important goal of STEM education in the Big Data era (Börner et al., 2019; Bybee, 2010; NRC, 2012; Pepler et al., 2021). This study investigated learners' difficulties and metacognitive patterns as they engaged in the practice of data visualization construction so that curricula and interventions can be developed to address learners' needs. Moreover, the study contributes to extending and elaborating the current understanding of the CMDC and metavisualization perspectives by providing concrete examples and empirical evidence of metacognitive processes in the context of data visualization. The study also provides a case of applying these perspectives for the topic of data visualization in STEM education.

This study identified six types of difficulties associated with students constructing data visualizations. The difficulties indicate areas that data science curricula or interventions need to consider in order to develop students' DVL, including the content knowledge to comprehend the information in the dataset and context, the technological content knowledge to know about the computer-based data visualization tool, and the knowledge and skills to simplify the data set and represent the data in visual forms.

Combining the qualitative coding and analysis and quantitative ENA technique, it can be concluded that the majority (about three-quarters) of the students were able to generate and use strategies to overcome the data visualization difficulties. The most frequent strategies used by

the students are the inducting and resourcing strategies to address the most common difficulties, including the technological and content difficulties. The study further distinguished the strategies into non-, basic- and high-level metavisual strategies, which provides insights into trajectories for supporting students in forming and using data visualization strategies.

Moreover, the different patterns of the college and high school students were identified in this study. Specifically, the high school students demonstrated more fundamental needs including the needs for content knowledge and representation skills. In comparison, the college students demonstrated the need for knowledge and skills regarding how to simplify the data to construct the best data visualization in a computer-based data visualization environment. Future research is needed to design data science curricula to address the different needs of the college and high school students and to investigate the effects.

One limitation of this study stems from the nature of the study which engaged participants in individual interview sessions; this is a trade-off for the purpose of focusing on individuals' data visualization processes. Future research may engage a group or class of students in a session and investigate the collaborative and interactive processes of data visualization among peers and experienced others to understand the social factors relating to data visualization. The affective factors were not investigated in this study. Moreover, the participants, the data visualization tool used and the visualization task implemented in the context of the study addressed the issue of developing novices' DVL. The results may not be generalizable to students with statistics or data science majors. Another limitation relates to the use of the think-aloud technique which has constraints depending on individuals' verbalization abilities, as well as whether metacognitive or cognitive processes are accessible (Jääskeläinen, 2010). Future work may combine other techniques such as eye tracking to obtain visual attention data for unconscious cognitive or metacognitive performance (e.g., Martinez et al., 2021).

Also, due to the voluntary nature of this study, which mainly attracted female students, the generalization of the results may be limited. Gender difference was not investigated here, but may be investigated in future studies, given that gender-based self-beliefs, self-regulation and metacognitive strategies are presently major topics in STEM education and research, and that little research has investigated gender differences in the topic of data science education. We found only one study investigating this topic, which reported that male students exhibited higher levels of learning motivation and expectancy for success in taking data science

courses, but no gender difference was observed in academic performance (Ivaniushina et al., 2016). While Ivaniushina et al.'s (2016) study indicated that male students exhibited higher motivation levels in taking data science courses, our study revealed another trend that is worthy of discussion and warrants further investigation. More female students willingly participated in the data visualization tasks and interviews in the study. However, the student population at the recruiting schools showed a nearly equal distribution of gender. Future research is needed to investigate the reason for the observed gender difference in this study. Nevertheless, we conjecture the possible significance of contextual factors. For instance, participation in research scenarios such as the visualization tasks and interviews in this study did not require the long-term commitment needed in the context of taking a data science course. How different contexts of learning data science may motivate different populations merits increased attention and warrants further exploration in future research. Finally, the study did not focus on reporting the quality of the visualization products or explore the relationships between the students' demonstrated strategies and difficulties and the quality of the final products, which can be continuing and future work.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40594-024-00463-w>.

Additional file 1: Codes, definitions and example excerpts for the strategies and difficulties (the excerpts were originally in Chinese and translated into English).

Acknowledgements

We would like to thank the researchers and research assistants who assisted the data collection, management, transcription and analysis procedure of the study.

Author contributions

H-YC: conceptualization, methodology, formal analysis, writing, review and editing; Y-JC: conceptualization, methodology, formal analysis, review and editing; M-JT: methodology, formal analysis, review and editing.

Funding

This work was supported by National Science and Technology Council, Taiwan, under Grants NSTC108-2628-H-003-001-MY3, NSTC111-2628-H-003-004-MY3, and MOST109-2511-H-003-018-MY3. It was also financially supported by the Institute for Research Excellence in Learning Sciences of National Taiwan Normal University (NTNU) from the Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education (MOE) in Taiwan.

Availability of data and materials

The data supporting the findings will be available upon request.

Declarations

Competing interests

We, as the authors of this study, declare that we have no competing interest.

Author details

¹Program of Learning Sciences and Institute for Research Excellence in Learning Sciences, National Taiwan Normal University, #162, Sec.1, Heping E. Rd., Taipei City 106, Taiwan. ²Department of Graphic Arts and Communications, National Taiwan Normal University, Taipei City, Taiwan.

Received: 17 May 2023 Accepted: 30 December 2023

Published online: 14 February 2024

References

- Angra, A., & Gardner, S. M. (2017). Reflecting on graphs: Attributes of graph choice and construction practices in biology. *CBE Life Sciences Education*, 16(3), ar53. <https://doi.org/10.1187/cbe.16-08-0245>
- Bae, S. S., Vanukuru, R., Yang, R., Gyory, P., Zhou, R., Do, E.Y.-L., & Szafr, D. A. (2023). Cultivating visualization literacy for children through curiosity and play. *IEEE Transactions on Visualization and Computer Graphics*, 29(1), 257–267. <https://doi.org/10.1109/TVCG.2022.3209442>
- Bargagliotti, A., Franklin, C., Arnold, P., Gould, R., Johnson, S., Perez, L., & Spangler, D. (2020). *Pre-K-12 guidelines for assessment and instruction in statistics education II (GAISE II): A framework for statistics and data science education*. American Statistical Association and National Council of Teachers of Mathematics.
- Bertin, J. (2011). *Semiology of graphics: Diagrams, networks, maps*. USA: ESRI Press.
- Binali, T., Chang, C.-H., Chang, Y.-J., & Chang, H.-Y. (2022). High school and college students' graph-interpretation competence in scientific and daily contexts of data visualization. *Science & Education*. <https://doi.org/10.1007/s11191-022-00406-3>
- Blackford, K. A., Greenbaum, J. C., Redkar, N. S., Gaillard, N. T., Helix, M. R., & Baranger, A. M. (2023). Metacognitive regulation in organic chemistry students: How and why students use metacognitive strategies when predicting reactivity. *Chemistry Education Research and Practice*, 24, 828–851. <https://doi.org/10.1039/D2RP00208F>
- Börner, K., Bueckle, A., & Ginda, M. (2019). Data visualization literacy: Definitions, conceptual frameworks, exercises, and assessments. *Proceedings of the National Academy of Sciences of the United States of America*, 116, 1857–1864. <https://doi.org/10.1073/pnas.1807180116>
- Börner, K., Maltese, A., Balliet, R. N., & Heimlich, J. (2016). Investigating aspects of data visualization literacy using 20 information visualizations and 273 science museum visitors. *Information Visualization*, 15(3), 198–213. <https://doi.org/10.1177/1473871615594652>
- Bressler, D. M., Bodzin, A. M., Eagan, B., & Tabatabai, S. (2019). Using epistemic network analysis to examine discourse and scientific practice during a collaborative game. *Journal of Science Education and Technology*, 28, 553–566. <https://doi.org/10.1007/s10956-019-09786-8>
- Bybee, R. W. (2010). Advancing STEM education: A 2020 vision. *Technology and Engineering Teacher*, 70(1), 30–35.
- Byrd, V. L., & Dwenger, N. (2021). Activity worksheets for teaching and learning data visualization. *IEEE Computer Graphics and Applications*, 41(6), 25–36. <https://doi.org/10.1109/MCG.2021.3115396>
- Camm, J. D., McCray, G. E., & Roehm, M. L. (2023). More than just charts and graphs: What to teach in a data visualization course. *Decision Sciences Journal of Innovative Education*, 21(3), 112–122. <https://doi.org/10.1111/dsji.12282>
- Chang, H.-Y. (2022). Science teachers' and students' metavisualization in scientific modeling. *Science Education*, 106(2), 448–475. <https://doi.org/10.1002/sce.21693>
- Chang, H.-Y., & Tsai, C.-C. (2023). Epistemic network analysis of students' drawings to investigate their conceptions of science learning with technology. *Journal of Science Education and Technology*, 32, 267–283. <https://doi.org/10.1007/s10956-022-10026-9>
- Chrysantina, A., & Sæbø, J. I. (2019). Assessing user-designed dashboards: A case for developing data visualization competency. In P. Nielsen & H. C. Kimaro (Eds.), *Information and communication technologies for development: Strengthening southern-driven cooperation as a catalyst for ICT4D* (pp. 448–459). Springer.
- De Boer, H., Donker, A. S., Kostons, D. D. N. M., & van der Werf, G. P. C. (2018). Long-term effects of metacognitive strategy instruction on student

- academic performance: A meta-analysis. *Educational Research Review*, 24, 98–115. <https://doi.org/10.1016/j.edurev.2018.03.002>
- diSessa, A. A. (2004). Meta-representation: Native competence and targets for instruction. *Cognition and Instruction*, 22, 293–331. https://doi.org/10.1207/s1532690xc12203_2
- Donnelly-Hermosillo, D. F., Gerard, L. F., & Linn, M. C. (2020). Impact of graph technologies in K-12 science and mathematics education. *Computers & Education*, 146, 103748. <https://doi.org/10.1016/j.compedu.2019.103748>
- Donohoe, D., & Costello, E. (2020). Data visualisation literacy in higher education: An exploratory study of understanding of a learning dashboard tool. *International Journal of Emerging Technologies in Learning*, 15(17), 115–126. <https://doi.org/10.3991/ijet.v15i17.15041>
- Doorman, M., Drijvers, P., Gravemeijer, K., Boon, P., & Reed, H. (2012). Tool use and the development of the function concept: From repeated calculations to functional thinking. *International Journal of Science and Mathematics Education*, 10, 1243–1267. <https://doi.org/10.1007/s10763-012-9329-0>
- Gilbert, J. K. (2005). Visualization: A metacognitive skill in science and science education. In J. K. Gilbert (Ed.), *Visualization in science education* (pp. 9–27). Springer.
- Gilbert, J. K. (2008). Visualization: An emergent field of practice and enquiry in science education. In J. K. Gilbert, M. Reiner, & M. Nakhleh (Eds.), *Visualization: Theory and practice in science education* (Vol. 3, pp. 3–24). Springer.
- Gilbert, J. K. (2010). The role of visual representations in the learning and teaching of science: An introduction. *Asia-Pacific Forum on Science Learning and Teaching*, 11(1), 1–19.
- Graham, C. R. (2011). Theoretical considerations for understanding technological pedagogical content knowledge (TPACK). *Computers & Education*, 57(3), 1953–1960. <https://doi.org/10.1016/j.compedu.2011.04.010>
- Grammel, L., Tory, M., & Storey, M.-A. (2010). How information visualization novices construct visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 16(6), 943–952. <https://doi.org/10.1109/TVCG.2010.164>
- Griese, B., Lehmann, M., & Roesken-Winter, B. (2015). Refining questionnaire-based assessment of STEM students' learning strategies. *International Journal of STEM Education*, 2, 12. <https://doi.org/10.1186/s40594-015-0025-9>
- Heller, J., Steiner, C., Hockemeyer, C., & Albert, D. (2006). Competence-based knowledge structures for personalised learning. *International Journal on E-Learning*, 5(1), 75–88.
- Hung, J.-Y., Chang, H.-Y., & Hung, J.-F. (2021). An experienced science teacher's metavisualization in the case of the complex system of carbon cycling. *Research in Science Education*, 51, 493–521. <https://doi.org/10.1007/s11165-018-9804-x>
- Ivaniushina, V., Alexandrov, D., & Musabirov, I. (2016). The structure of students' motivation: Expectancies and values in taking data science course. *Voprosy Obrazovaniya Educational Studies Moscow*, 4, 229–250. <https://doi.org/10.17323/1814-9545-2016-4-229-250>
- Jääskeläinen, R. (2010). Think-aloud protocol. In Y. Gambier & L. van Doorslaer (Eds.), *Handbook of translation studies* (Vol. 1, pp. 371–373). USA: John Benjamins Publishing Company. <https://doi.org/10.1075/hts.1.th11>
- Justi, R., Gilbert, J. K., & Ferreira, P. F. (2009). The application of a model of modelling to illustrate the importance of metavisualisation in respect of the three types of representation. In J. K. Gilbert & D. Treagust (Eds.), *Multiple representations in chemical education* (pp. 285–307). Springer.
- Kaliisa, R., Misiejuk, K., Irgens, G. A., & Misfeldt, M. (2021). Scoping the emerging field of quantitative ethnography: Opportunities, challenges and future directions. In A. R. Ruis & S. B. Lee (Eds.), *Advances in quantitative ethnography* (pp. 3–17). Springer. https://doi.org/10.1007/978-3-030-67788-6_1
- Karlen, Y., Merki, K. M., & Ramseier, E. (2014). The effect of individual differences in the development of metacognitive strategy knowledge. *Instructional Science*, 42, 777–794. <https://doi.org/10.1007/s11251-014-9314-9>
- Ku, K. Y. L., & Ho, I. T. (2010). Metacognitive strategies that enhance critical thinking. *Metacognition and Learning*, 5, 251–267. <https://doi.org/10.1007/s11409-010-9060-6>
- Lee, S., Kim, S. H., & Kwon, B. C. (2017). VLAT: Development of a visualization literacy assessment test. *IEEE Transactions on Visualization and Computer Graphics*, 23(1), 551–560. <https://doi.org/10.1109/tvcg.2016.2598920>
- Lee, S., Kwon, B. C., Yang, J., Lee, B. C., & Kim, S.-H. (2019). The correlation between users' cognitive characteristics and visualization literacy. *Applied Sciences*, 9(3), 488. <https://doi.org/10.3390/app9030488>
- Li, Q. (2020). Overview of data visualization. In Q. Li (Ed.), *Embodying data* (pp. 17–47). Springer. https://doi.org/10.1007/978-981-15-5069-0_2
- Linn, M. C., Eylon, B.-S., & Davis, E. A. (2004). The knowledge integration perspective on learning. In M. C. Linn, E. A. Davis, & P. Bell (Eds.), *Internet environments for science education* (pp. 29–46). Routledge.
- Locatelli, S., & Arroio, A. (2014). Metavisual strategy assisting the learning of initial concepts of electrochemistry. *Gamtamokslinis Ugdymas Natural Science Education*, 11(1), 14–24.
- Locatelli, S., & Arroio, A. (2016). Contributions and limitations of a metavisual strategy from the perspective of students. In J. Lavonen, K. Juuti, J. Lampiselkä, A. Uitto, & K. Hahl (Eds.), *Electronic proceedings of the ESERA 2015 Conference. Science education research: Engaging learners for a sustainable future, part 3/3 (Sabine Fechner & Andrée Tiberghien)* (pp. 437–442). University of Helsinki.
- Locatelli, S., & Davidowitz, B. (2021). Using metavisualization to revise an explanatory model regarding a chemical reaction between ions. *Chemistry Education Research and Practice*, 22(2), 382–395. <https://doi.org/10.1039/D0RP00339E>
- Mansoor, H., & Harrison, L. (2018). Data visualization literacy and visualization biases: Cases for merging parallel threads. In G. Ellis (Ed.), *Cognitive biases in visualizations* (pp. 87–96). Springer.
- Marquart, C. L., Hinojosa, C., Swiecki, Z., Eagan, B., & Shaffer, D. W. (2018). *Epistemic Network Analysis (Version 1.5.2) [Software]*
- Martinez, B. L., Sweeder, R. D., VandenPlas, J. R., & Herrington, D. G. (2021). Improving conceptual understanding of gas behavior through the use of screencasts and simulations. *International Journal of STEM Education*, 8, 5. <https://doi.org/10.1186/s40594-020-00261-0>
- Merisio, C., Bozzi, G., & Datteri, E. (2021). There is no such thing as a "trial and error strategy." In M. Malvezzi, D. Alimisis, & M. Moro (Eds.), *Education in & with robotics to foster 21st-century skills: Proceedings of EDUROBOTICS 2020* (pp. 190–201). Springer.
- National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. The National Academies Press.
- Ow-Yeong, Y. K., Yeter, I. H., & Ali, F. (2023). Learning data science in elementary school mathematics: a comparative curriculum analysis. *International Journal of STEM Education*, 10, 8. <https://doi.org/10.1186/s40594-023-00397-9>
- Peppler, K., Keune, A., & Han, A. (2021). Cultivating data visualization literacy in museums. *Information and Learning Sciences*, 122(1/2), 1–16. <https://doi.org/10.1108/ILS-04-2020-0132>
- Rachmatullah, A., & Wiebe, E. N. (2022). Building a computational model of food webs: Impacts on middle school students' computational and systems thinking skills. *Journal of Research in Science Teaching*, 59(4), 585–618. <https://doi.org/10.1002/tea.21738>
- Ruis, A. R., & Lee, S. B. (2021). Advances in quantitative ethnography. Springer. <https://doi.org/10.1007/978-3-030-67788-6>
- Schraw, G., & Moshman, D. (1995). Metacognitive theories. *Educational Psychology Review*, 7, 351–371. <https://doi.org/10.1007/BF02212307>
- Shaffer, D. W. (2014). *Formatting data for epistemic network analysis* (Technical Report 2014–1). Epistemic Games Group.
- Shaffer, D. W. (2017). *Quantitative ethnography*. Cathcart Press.
- Shaffer, D. W. (2018). Epistemic network analysis: Understanding learning by using big data for thick description. In F. Fischer, C. E. Hmelo-Silver, S. R. Goldman, & P. Reimann (Eds.), *International handbook of the learning sciences* (1st ed., pp. 520–531). Routledge.
- Shaffer, D. W., Collier, W., & Ruis, A. R. (2016). A tutorial on epistemic network analysis: Analyzing the structure of connections in cognitive, social, and interaction data. *Journal of Learning Analytics*, 3(3), 9–45.
- Steiner, C. M., & Albert, D. (2011). Competence-based knowledge space theory as a framework for intelligent metacognitive scaffolding. In G. Biswas, S. Bull, J. Kay, & A. Mitrovic (Eds.), *Artificial intelligence in education: 15th International conference, AIED 2011* (Vol. 6738, pp. 563–565). Springer. https://doi.org/10.1007/978-3-642-21869-9_102
- Sun, M., Wang, M., Wegerif, R., & Peng, J. (2022). How do students generate ideas together in scientific creativity tasks through computer-based mind mapping? *Computers & Education*, 176, 104359. <https://doi.org/10.1016/j.compedu.2021.104359>
- Unwin, A. (2020). Why is data visualization important? What is important in data visualization? *Harvard Data Science Review*, 2(1). <https://doi.org/10.1162/99608f92.8ae4d525>

- van Dijk, T. A. (1981). Episodes as units of discourse analysis. In D. Tannen (Ed.), *Analyzing discourse: Text and talk* (pp. 177–195). Georgetown University Press.
- Van Meter, P., & Firetto, C. M. (2013). Cognitive model of drawing construction: Learning through the construction of drawings. In G. J. Schraw, M. T. McCrudden, & D. R. Robinson (Eds.), *Learning through visual displays* (pp. 247–280). Information Age Publishing.
- Wainer, H. (1992). Understanding graphs and tables. *Educational Researcher*, 21, 14–23. <https://doi.org/10.3102/0013189X021001014>
- Wooldridge, A. R., Carayon, P., Shaffer, D. W., & Eagan, B. (2018). Quantifying the qualitative with epistemic network analysis: A human factors case study of task-allocation communication in a primary care team. *IJSE Transactions on Healthcare Systems Engineering*, 8(1), 72–82.
- Zhang, S., Gao, Q., Sun, M., Cai, Z., Li, H., Tang, Y., & Liu, Q. (2022). Understanding student teachers' collaborative problem solving: Insights from an epistemic network analysis (ENA). *Computers & Education*, 183, 104485. <https://doi.org/10.1016/j.compedu.2022.104485>
- Zörgő, S., Swiecki, Z., & Ruis, A. R. (2021). Exploring the effects of segmentation on semi-structured interview data with epistemic network analysis. In A. R. Ruis & S. B. Lee (Eds.), *Advances in quantitative ethnography* (pp. 78–90). Springer. https://doi.org/10.1007/978-3-030-67788-6_6

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ [springeropen.com](https://www.springeropen.com)
