Reflecting on Graphs: Attributes of Graph Choice and Construction Practices in Biology

Aakanksha Angra† and Stephanie M. Gardner‡*

†School of Biological Sciences, Georgia Institute of Technology, Atlanta, GA 30332; ‡Department of Biological Sciences, Purdue University, West Lafayette, IN 47907

ABSTRACT

Undergraduate biology education reform aims to engage students in scientific practices such as experimental design, experimentation, and data analysis and communication. Graphs are ubiquitous in the biological sciences, and creating effective graphical representations involves quantitative and disciplinary concepts and skills. Past studies document student difficulties with graphing within the contexts of classroom or national assessments without evaluating student reasoning. Operating under the metarepresentational competence framework, we conducted think-aloud interviews to reveal differences in reasoning and graph quality between undergraduate biology students, graduate students, and professors in a pen-and-paper graphing task. All professors planned and thought about data before graph construction. When reflecting on their graphs, professors and graduate students focused on the function of graphs and experimental design, while most undergraduate students relied on intuition and data provided in the task. Most undergraduate students meticulously plotted all data with scaled axes, while professors and some graduate students transformed the data, aligned the graph with the research question, and reflected on statistics and sample size. Differences in reasoning and approaches taken in graph choice and construction corroborate and extend previous findings and provide rich targets for undergraduate and graduate instruction.

INTRODUCTION

Graphs are the main components of the scientific language, because they can be used to condense and summarize large data sets. The result is a symbolic representation that displays experimental findings used by scientists for communication (Beichner, 1994; Tairab and Al-Naqbi, 2004; Wainer, 2013). The development of the skill to create appropriate and clear graphs is necessary for the scientifically literate individual (Padilla *et al.*, 1986). Indeed, recent calls to reform the undergraduate curriculum include incorporating aspects of data literacy into the science, technology, engineering, and mathematics disciplines. Within the discipline of biology, there is an emphasis on the infusion of quantitative reasoning into the classroom, including creating and interpreting graphical representations (Association of American Medical Colleges, 2009; American Association for the Advancement of Science, 2011). The increasing implementation of course-based undergraduate research experiences (CUREs) emphasizes the importance of understanding how students grapple with data and data presentation to facilitate their mastery of this skill (see Figure 1 in Auchincloss *et al.*, 2014). Furthermore, current studies in the field of biology education have shown that students who engage in research practices feel more inclusive in the learning process and gain better science process skills, such as data analysis and graphing (Bangera and Brownell, 2014; Brownell *et al.*, 2015; Linn *et al.*, 2015).

CONCEPTS AND SKILLS NEEDED FOR GRAPHING AND AREAS OF DIFFICULTY

The purpose of a graph is to communicate observational or numerical data in a visual format (Tufte, 1983; Leinhardt *et al.*, 1990), with the hope that the graph is interpreted in the same manner and with the same take-home message as the graph constructor intended. Extensive research has documented student difficulties with graph interpretation. Tairab and Al-Naqbi (2004) showed that students in 10th grade had difficulty understanding that the *x*- and *y*-axes illustrate the relationship between the independent and dependent variables. Other studies show similar difficulties with interpreting interactions and slope of a line (Preece and Janvier, 1992; Picone *et al.*, 2007; Colon-Berlingeri and Burrowes, 2011).

While these studies focused on graph interpretation, the concepts and skills that they studied are integral to graph construction as well. Before constructing the graph, the graph constructor should have a clear purpose in mind, along with an adequate understanding of variables and graph types (Berg and Smith, 1994; Friel and Bright, 1996; Clase *et al.*, 2010; Grunwald and Hartman, 2010; Angra and Gardner, 2016). For a graph to be an effective communication piece for both the creator and the observer, four main components should be considered: 1) data form, 2) graph choice, 3) graph mechanics, and 4) aesthetics and visuospatial aspects. While these are four distinct components, they are all interrelated and influence the type and quality of the message communicated by the graph (Table 1). For example, the form of the plotted data (e.g., raw data vs. averages) can influence the type of graph and labeling used to clearly display those data.

Owing to its complexity, choosing and constructing an appropriate graph for data can be considered a problem-solving task (Angra and Gardner, 2016). Our previous findings (see Figure 1 in Angra and Gardner, 2016) on the steps taken during a pen-and-paper graphing construction task by expert professors resembled the four steps of Polya's problem-solving cycle in mathematics (Polya, 1945). Polya's problem-solving model has been adapted based on the data and trends that have emerged from our work to explain expert graph-construction behavior and can be distilled into three phases: planning, execution, and reflection (for a detailed description, see Angra and Gardner, 2016). During the planning phase, before the graph is constructed, data to be plotted are evaluated, understood, and characterized. Specifically, decisions on the purpose for graphically displaying the data are clarified, ways to organize the data on the graph are considered, decisions on data transformation are made, and a graph type is chosen (Friel and Bright, 1996; Ainley *et al.*, 2000; Patterson and Leonard, 2005; Angra and Gardner, 2016). During the execution phase, the graph is constructed with appropriate elements of graph mechanics for clear communication (e.g., descriptive title, variables on axes, scales appropriate for data, key, etc.) and data are plotted (Angra and Gardner, 2016). Finally, during the reflection phase, the constructed graph is critiqued, graph choice is evaluated, and the graph is checked for alignment with the intended purpose (Angra and Gardner, 2016).

As noted, current trends in biology education engage students in data analysis and graphing; however, students across the K–16 continuum struggle with many fundamental concepts and skills relevant for graphing, including scaling axes, using a best-fit line, and assigning variables to axes (Padilla *et al.*, 1986). Further, while there are standards and recommendations for K–16 education in areas related to quantitative literacy (Aliaga *et al.*, 2005), standards for graduate education have been lacking. There have been increased efforts to formalize quality training for graduate students as instructors

TABLE 1. Criteria for evaluating graph attributes explaining the components of graph mechanics, data form, graph choice, and aesthetics

Categories used to describe graphs qualitatively	Category descriptions	Citations
Graph mechanics	1. Title: a title should be descriptive for the graph. 2. Axes labels: both the *x*- and *y*-axis labels should be appropriate and descriptive for the experiment. 3. Units: should be appropriate and descriptive for the type of data displayed. 4. Scale: should be appropriate for the data displayed such that the increments are clear and easy to understand.	Padilla *et al.*, 1986; Li and Shen, 1992; Brasell and Rowe, 1993; Kosslyn, 1994; Kostelnick, 1998; Ainley, 2000; Konold and Higgins, 2003; Leonard and Patterson, 2004; Bruno and Espinel, 2009; Bray-Speth *et al.*, 2010; McFarland, 2010
Data form	1. Graph should show a clear distinction between raw and manipulated data plotted.	Wild and Pfannkuch, 1999; Konold *et al.*, 2015
Graph choice	1. Graph type: graph type should be appropriate for both the independent and dependent variables. 2. Alignment: graph should align with the original intended purpose. 3. Take-home message: graph type allows reader to draw appropriate conclusions from the data in the graph.	Cleveland, 1984; Li and Shen, 1992; Bright and Friel, 1998; Shah *et al.*, 1999; Schriger and Cooper, 2001; Konold and Higgins, 2003; Grawemeyer and Cox, 2004; Leonard and Patterson, 2004; Metz, 2008; Bray-Speth *et al.*, 2010; McFarland, 2010; Franzblau and Chung, 2012; Humphrey *et al.*, 2014; Rougier *et al.*, 2014; Slutsky, 2014; Angra and Gardner, 2016
Aesthetics and visuospatial aspects	1. The graph should be pleasing to the eye such that the data plotted occupy sufficient room in the Cartesian plane. 2. Sound construction and mechanistic properties enable the reader to extract meaning from the graph.	Tufte, 1983; Kosslyn, 1994; Kostelnick, 1998; Kellman, 2000; Few, 2004

(Schussler *et al.*, 2008; Reeves *et al.*, 2016) and scholars (National Institutes of Health [NIH], 2016; National Science Foundation [NSF], 2016). However, specific objectives for concepts and skills for all graduate students to master have not been widely implemented outside the activities of funded programs, such as training grants (NIH, 2016). Quantitative skills related to data representation are most likely developed by graduate students through experience reading primary literature, analyzing and presenting their own data, and with guidance from their research mentors. However, graphing difficulties exist and have been documented in individuals who possess advanced and/or terminal degrees, that is, professors (Bowen and Roth, 2005), professionals (Rougier *et al.*, 2014; Weissgerber *et al.*, 2015) and medical doctors (Cooper *et al.*, 2001, 2002; Schriger and Cooper, 2001; Schriger *et al.*, 2006).

Previous studies share suggestions and sample data sets to encourage practice with graph creation (Tairab and Al-Naqbi, 2004; Patterson and Leonard, 2005; Bray-Speth *et al.*, 2010). For instance, Patterson and Leonard (2005) advocate for training students to use software for graph construction, using a balance of analytical thought and creative artistry. However, before letting students use software, they suggest that students should focus on the message they want to communicate in a graph, explain the appropriate statistics, and sketch a graph by hand so they know what the end product produced by the software should look like (Patterson and Leonard, 2005). Other suggestions to remediate graphing difficulties include incorporating graphing into the science classroom. This will provide more opportunities, repetition, and student–instructor feedback to tackle graphing difficulties and increase student competency with graphing (Roth and McGinn, 1997; Roth and Bowen, 2001; McFarland, 2010; Harsh and Schmitt-Harsh, 2016).

The best methods and techniques for graph construction when translating raw data into a graph are still unknown, which can lead to challenges for both undergraduate and graduate stu-

dents and active research scientists. The underlying thought processes used by graph constructors when choosing and constructing graphs are not fully understood. Therefore, one problem we face is having an incomplete understanding of the reasoning that occurs during graph choice and construction. While constructing a graph using software programs is useful and replicates the authentic graph-making processes that occur in classrooms and laboratories, it can interfere with thoughtful and reflective decision making. Software programs overload the graph constructor with multiple graphing choices, without having the graph constructor reflect on decisions regarding variables, data, graph choice, and the purpose of the graph. In this study, we aim to uncover the reasoning that occurs during graph choice and construction and the attributes of the resulting graphs by using the pen-and-paper mode of graph construction.

THEORETICAL FRAMEWORK GUIDING STUDY DESIGN AND ANALYSIS

Our study design and data analysis are guided by the metarepresentational competence (MRC) framework (diSessa and Sherin, 2000). This framework outlines the knowledge and reflective reasoning practices that an individual competent in creating external representations (e.g., graphs), such as an expert scientist, would exhibit. As such, implicit in the MRC framework are expert-like knowledge and skill (diSessa, 2004), which can provide helpful benchmarks when studying student MRC (National Research Council, 2000; diSessa, 2004) and can inform classroom practices. The components of the MRC framework can be leveraged to reveal a person's areas of competence and difficulty with graph choice, construction, and critique. Specifically, these components are invention, critique, functioning, and learning or reflection (diSessa and Sherin, 2000; summarized in Table 2). In our study the MRC component of invention is assumed, because all participants created a graph. Therefore, we use the last three components from MRC to

TABLE 2. Categories in the MRC and their definitions and connections to this study

Categories in the MRC	Definitions[a]	Connection to this study
Invention	The underlying skills and abilities that allow students to conceive novel representations	• Competency with graph choice, construction, and knowledge of variables is vital for conjuring new graphical representations (Leonard and Patterson, 2004; Tairab and Al-Naqbi, 2004; McFarland, 2010; Webber *et al.*, 2014). • In the think-aloud interviews, participants were asked to construct a graph from raw data.
Critique	Critical knowledge that is essential for assessing the quality of representations	• Assessing the strengths and weaknesses of various graphs exposes students' critical knowledge (McFarland, 2010; Angra and Gardner, unpublished data) • Although the interviewer did not explicitly probe the participants to critique their graphs, we wanted to see whether participants spontaneously generated a critique.
Functioning	Providing reasoning for understanding the purpose of different representations, their usage, and limitations	• Functioning unearths students' reasoning for understanding the purpose of different types of graphs and the usage being dependent on the type of data present (McFarland, 2010; Webber *et al.* 2014; Angra and Gardner, unpublished data) • In the think-aloud interviews, students were asked to articulate their graph choice.
Learning/ reflection	Strategies for fostering understanding of representations	• Reflection, reveals students' awareness of their own understanding of graphs and gaps in their knowledge (Tanner, 2012) • Several times in the think-aloud interviews, participants were probed to reflect on their graph choice and construction.

[a]See diSessa and Sherin (2000).

define graph-construction reasoning as a persons' reflection on graph choice and construction by understanding the function of different types of graphs and being able to thoughtfully analyze a graph based on the type of data it is representing, variables, and the overall advantages and disadvantages of the chosen graph. As diSessa (2004) argues, creating a graph is not a difficult task, but the act of being critical, reflecting on the task and the graph itself, is what needs to be practiced to gain automaticity and independence with graphing.

RESEARCH QUESTIONS

The overarching research objective of this study is to elucidate the differences in graph-construction reasoning that may exist among undergraduate students, graduate students, and professors in the biological sciences. To accomplish this objective, we sought to answer two questions:

1. How do undergraduate students, graduate students, and professors reason with graph choice, data, and graph construction?
2. How do graph attributes differ between undergraduate students, graduate students, and professors?

METHODOLOGY

Think-Aloud Interviews for Graph Construction

In this study, we used a pen-and-paper graphing task in the context of think-aloud interviews to describe the reasoning behind graph choice and construction and the final graph artifacts. All interviews were conducted between March 2013 and October 2014. The LiveScribe pen was used to collect data, as it synchronizes written notes with recorded audio and has an embedded infrared camera that detects pen strokes when used with the LiveScribe dot paper (LiveScribe, 2015). Participants were randomly presented one of two scenarios (i.e., bacteria or plant scenario; Supplemental Material, Table 1) predetermined before the interview. Participants were asked to read the scenario prompt aloud and were then instructed to create a graph from the data in the scenario, narrating their thought process during this graph-construction task. Constructing a graph by hand may not be an everyday activity that most participants engage in, neither is thinking aloud while performing a task. To account for this, the interviewer gently probed the participants to articulate their thinking, especially if there were prolonged silences during graph construction. The think-aloud format provided insight into the thought process and reasoning, which was then used to characterize and delineate differences between experts and novices (Angra and Gardner, 2016). Think-aloud interviews are reliable sources of data, because they reveal the thought processes that occur and the sequences of thought (Ericsson, 2006). Several studies have found no evidence for differences in the accuracy of performance between those who silently completed the task versus those who verbalized their thoughts (Ericsson and Simon, 1993; Ali and Peebles, 2011). This gave us confidence that active narration would not influence the performance with the graphing task. After the participants finished their graph construction, the interviewer intervened and asked them to reflect on the following questions:

1. Why did you decide to create the graph that you did?
2. What are you plotting (raw data, computed value, etc.)?

The graphing task, with associated interview, ranged between 10 and 30 minutes in duration.

Development of the Graphing Scenario

The development of the scenario used in our think-aloud interviews involved outside validation and literature review. Knowing that some of our participants would have had at most a partial semester of introductory biology at the time of the interview, we consulted an award-winning high school teacher to get her opinion on biological scenarios that would be familiar to students who had ninth-grade biology. We used two scenarios: bacterial growth or plant growth (Table 1 of the Supplemental Material), because we wanted to minimize the threats to internal validity: instrumentation and diffusion of treatment (Drost, 2011). Both bacteria and plant scenarios are isomorphic, consisting of a dependent variable, independent variable, and two treatments with three replicates in each treatment. Simple numbers were used, so participants could easily manipulate the data, if they chose to do so (Konold *et al.*, 2015). In four sentences, the scenario provided the participants with a brief background and a data table that organized the elements mentioned earlier. We organized data in a table instead of a paragraph with numbers, because in scientific practice, data are often initially organized in a table so that it is easy for the graph constructor to visualize the raw values (Wainer, 2013). To validate the graph-construction prompts, we piloted the plant and bacteria scenarios with two undergraduate biology students and one professor. Pilot interviews were conducted in Fall 2012 to solidify the interview protocol and prompts and gauge the amount of time it took to construct a graph (Seidman, 2013). To ensure that the graphing scenario and task of constructing a graph while thinking aloud aligned our research questions, pilot interviews were transcribed and memoed (Patton, 2001) to look for ideas previously reported in the graphing literature.

Participant Recruitment

As part of a larger, multipart graphing study, undergraduate students, graduate students, and professors were recruited from the biological sciences department at a large, midwestern research university. A stratified, purposeful sampling method was used to obtain the target population (Hatch, 2002). To obtain a heterogeneous and representative sample of the undergraduate student population, we sent recruitment emails to faculty teaching large biology courses. Personal recruitment emails were sent to graduate students and biology faculty from diverse biological subdisciplines. All recruitment methods were approved by the Institutional Review Board (protocol no. 1210012775). Recruitment criteria for undergraduate students were based on 1) their status as or intention to be a biology major and 2) their current enrollment in or successful completion of the introductory biology lecture and laboratory course. At the time of recruitment, undergraduate research experience was not one of our criteria, but it was incorporated postinterview, based on literature outlining data representation skills and concepts students learn while engaged in research (Auchincloss *et al.*, 2014). In this paper, we report data from undergraduate students who did not have research experience at the time of the interview (UGNRs) and undergraduate students who did have research experience (UGRs). Recruitment

criteria for graduate students (GSs) were based on 1) their enrollment in the graduate program—all graduate students were pursuing a PhD degree; 2) successful completion of their qualifier examination taken at the end of their first year; and 3) their having held a teaching assistantship or having mentored undergraduate students. Criteria for professors were based on 1) their credentials—all professors held a PhD in a subdiscipline of biology; 2) their having an active research laboratory with postdocs, graduate students, and/or undergraduate students; and 3) their having taught for at least 1 year.

Participants and Inclusion Criteria

Our initial pool of participants included seven professors, 13 graduate students, and 39 undergraduate students. This pool was reduced based on the following inclusion criteria. To minimize the threat to internal validity, we eliminated the six undergraduate and one graduate student interviews that were conducted early in the project with an interviewer who did not follow the think-aloud protocol with high fidelity. From the remaining 33 undergraduate student interviews that were conducted by the first author (A.A.), we further eliminated students who spontaneously constructed multiple graphs during the first prompt to construct a graph, as they did not articulate their reflection on graph choice for all graphs they constructed, and the interviewer felt it was inappropriate to interrupt the flow of thought during graph construction. Although these data are interesting and will be analyzed in future work, for this study, we chose to exclude them to ensure uniformity across all participant groups. The same criteria were applied to graduate students and professors. Our final participant pool consisted of five professors, eight graduate students, and 15 undergraduate students. Of the 15 undergraduate students, 10 reported having no research experience and five reported having research experience. In this study, we categorized and defined our most novice participants as the ones who reported not having any research experience, followed by undergraduate students who reported research experience, graduate students, and finally, the professors, who each had more than 10 years' experience conducting research and constructing graphs. Participants in our study represented many subdisciplines in biology. Professors' specialties ranged from cellular neurobiology to behavioral ecology, while the graduate students' research interests ranged from virology to avian behavior. The Supplemental Material, Tables 2–4, provides demographic information for our participants. Because undergraduate research experiences vary immensely, we found that using the relative approach described here to group experts as professors, graduate students as advanced, undergraduate students with research experience as intermediates, and undergraduate students without research experience as novices (Chi, 2006) to be a useful method of analysis.

DATA ANALYSIS

Data Organization and Coding

Think-aloud interviews were transcribed verbatim and systematically organized and coded using inductive analysis to address the first research question (Strauss and Corbin, 1998; Patton, 2001). This initial step of transcript segmentation began the process of open coding within each phase of thought (planning, execution, and reflection phases). Selective coding was then used to organize the codes into a story that described the complex network of themes that emerged (Creswell, 2013). For the final step, themes from the selective coding step were aligned to the categories present in the MRC framework. The first author (A.A.) independently coded all transcripts from the think-aloud interviews and compared her codes with 20% of those coded by the second author (S.M.G.). Both authors met regularly to compare and discuss the coding, until a consensus was reached on the final codes and themes.

To see whether there was a difference among the participant groups in terms of the time it took to plan, construct, and reflect on the graph, we conducted an independent-samples t test using Statistical Package for the Social Sciences, version 22 (SPSS v. 22; IBM, 2013). Levene's test for the equality of variance was conducted, and equal variances were not assumed when reporting the p value ($\alpha < 0.05$; SPSS v. 22; IBM, 2013). Because we were interested in differences across participant groups, we did not perform inferential statistics across phases of the graph interview. Professors also used more words than undergraduate students in their thought processes and explanations. Roth and Bowen (2003) used word analysis to understand how experts interpreted graphs. We used a similar method to quantify and characterize the number of words spoken during each phase by the participants. Transcripts were coded in Microsoft Word by placing portions of the interview transcript under specific codes in our codebook. To standardize time spent talking by each participant, we performed word analysis. Words mentioned multiple times within a given phase were counted and coded once. The words for each code were counted and the number was divided by the total number of words uttered by the participant. This number was then multiplied by 100 to obtain the percentage of words uttered for particular codes for the particular phase. Final results are displayed in Figure 1.

Owing to the small sample size in each participant group, statistics for themes on the qualitative interview data are not reported, but the absence or presence of themes and the occurrence of the MRC categories between the three participant groups are summarized in Figure 2.

For addressing the second research question, graphs constructed by professors, GSs, and the two undergraduate population groups (UGNR and UGR) were described qualitatively based on four broad categories: graph mechanics, data form, graph choice, and aesthetics. The evaluation categories are listed in Table 1.

RESULTS

Graph-Construction Reasoning

To answer our first research question, we identified the themes that emerged from the transcripts from our think-aloud graph-construction interviews for each phase of the graph-construction process (planning, construction, and reflection). We mapped the emergent themes to the categories of the MRC framework.

Planning Phase

The planning phase occurred after participants were presented with the task and before they began graph construction, as indicated by the drawing of the axes. Figure 1 displays the amount of time the participants spent talking in each of the interview
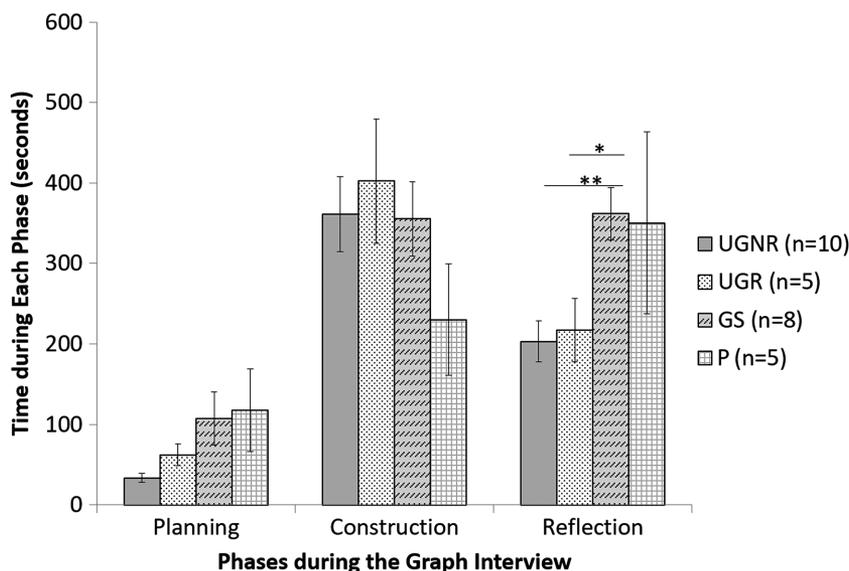
FIGURE 1. A comparison of the amount of time spent in each interview phase by undergraduate students, graduate students, and professors that summarizes the time spent during the planning, construction, and reflection phases for professors (P, $n = 5$), graduate students (GS, $n = 8$), undergraduates with research experience (UGR, $n = 5$), and undergraduates without research experience (UGNR, $n = 10$). An independent-samples t test shows there was a significant difference in the amount of time spent reflecting between GS and UGR (*, $p < 0.05$) and GS and UGNR (**, $p < 0.01$).

phases. Looking across the three phases and at the four participant groups, we notice that, relative to the other two phases in the interview, participants spent the smallest amount of time planning. Within the planning phase, almost everyone took time to think about the scenario and data before proceeding with graph construction. This is indicated by the sample size in the second column in Figure 2.

Three out of the four categories from the MRC framework map onto the planning phase: function, invention, and learning/reflection (Figure 2 and Table 3). The definitions of the themes, example quotes, and the alignment of the themes to the MRC categories can be found in Table 3. Within the MRC category invention, the themes of data type and graph construction were prevalent across the multiple participant groups. However, the theme data type was seen only for one UGR, and the theme graph construction was seen for only one UGNR. Within the MRC category function, the themes purpose and graph choice emerged. In the planning phase, the theme purpose was observed only for

Components of MRC		INVENTION			CRITIQUE		FUNCTION				LEARNING/REFLECTION		
Themes from think-aloud interviews		Data Type	Statistics	Graph Construction	Sample Size	Aesthetics	Purpose	Variables	Time	Graph Choice	Technology	Data Table	Evaluation
Planning Phase	P (n=5)	<u>X</u>		<u>X</u>			<u>X</u>					<u>X</u>	
	GS (n=7)	<u>X</u>		<u>X</u>						<u>X</u>		<u>X</u>	
	UGR (n=4)	X		<u>X</u>			<u>X</u>			<u>X</u>		<u>X</u>	
	UGNR (n=7)	<u>X</u>		X						X		<u>X</u>	
Construction Phase	P (n=4)	<u>X</u>	<u>X</u>	<u>X</u>	<u>X</u>								<u>X</u>
	GS (n=7)	<u>X</u>	<u>X</u>	<u>X</u>						<u>X</u>	<u>X</u>		<u>X</u>
	UGR (n=5)	<u>X</u>	X	<u>X</u>		X				<u>X</u>	<u>X</u>		<u>X</u>
	UGNR (n=7)	<u>X</u>	<u>X</u>	<u>X</u>	X	<u>X</u>				<u>X</u>			<u>X</u>
Reflection Phase- Graph Choice	P (n=5)		X				<u>X</u>	X	<u>X</u>	<u>X</u>			<u>X</u>
	GS (n=8)		X			X	<u>X</u>	<u>X</u>	X	<u>X</u>		X	<u>X</u>
	UGR (n=5)		<u>X</u>			X			<u>X</u>	<u>X</u>		X	<u>X</u>
	UGNR (n=10)		<u>X</u>							<u>X</u>		<u>X</u>	<u>X</u>

FIGURE 2. Summary of graph-construction reasoning findings showing the presence of themes in each of the three interview phases by professors (P, $N = 5$), graduate students (GS, $N = 8$), undergraduates with research experience (UGR, $N = 5$), and undergraduates without research experience (UGNR, $N = 10$). "X" denotes the presence of a theme by one participant; "<u>X</u>" indicates the presence of a theme by multiple participants. Because invention involves graph construction and participants were explicitly asked to reflect on graph choice, these themes are blacked out. Refer to Figures 2, 3, and 4 for themes that appeared for each participant. Small n in the table is a subset of the total sample (N) of the participant group.

TABLE 3. Planning phase: summary of the themes, definitions, and participant examples

Categories in MRC	Themes	Participant examples
Function	Purpose: this is when the participant explicitly states that the purpose of the graph is to align with the purpose of the task.	P2: So the question is how temperature affects growth of bacteria. P4: We might be interested in taking a particular time point that we think is key and looking at the data for the groups at that time point, or we might sort of go the whole nine yards and [make] 5 separate plots. UGR2: Okay so we're measuring how one bacteria type grows at two different temperatures, so we have the two different temperatures and there are three tubes for each temperature, and we have different times so you can see how it grows.
	Graph choice: the participant is explicitly stating graph choice (i.e., bar, line, scatter) based on the data provided in the table. Participants may also interject their personal feelings or rely on their past experiences when contemplating between different graph types, their usage, and limitations.	GS4: I could make a scatter plot and have different symbols for different temperatures and they have three replicates for each. UGR1: Okay I'm going to make a bar graph for the sake of comparison here. Actually I might want to change my mind about what I'm doing here. I think I'm going to change to a type of line graph. I don't think the bars are going to be the best comparison for showing a time course of a single plant. UGNR4: I'm going to make a line graph to compare two different types of data in the same graph … I think it's going to show best patterns of each.
Invention	Graph construction: when the participant either verbalizes variables in the table to the axes on the graph or the data or explains how they are visualizing the data on the graph.	P1: So what I'm going to do is because the dependent measure is number of cells, I'm going to put that on the y axis. GS3: So definitely the independent variable is time, as they call it, the x axis. UGR1: So it will be plants with 15 ml of water per day and the bar with the lines will be for the 5 ml treatment group. UGNR9: So generally when you have time you want to put that on the x axis.
	Data type: the participant is explicitly making decisions about whether or not to plot raw data or plot manipulated data (i.e., average) or the number of graphs to use to properly convey the data.	P5: Let's start with the 15 ml [treatment]. What I would do is, since we have 3 points per time point I will try to get the average of the three. GS1: I would probably pull the replicates [together], although the math for this would be pretty bad in my head, and I would have to draw fake error bars. UGNR1: Well I might be able to make this into two graphs because it'll be easier to see maybe ... or we could do the average of the three tubes.
Learning or reflection	Data table: this is when the participant is making sense of the data provided in the data table as evidenced by summarizing the data and/or the variables presented.	P5: Number of leaves, 2 different amounts of water and you have time on axis and you have for each amount of water from plants. Ok. So now I should create a graph of that. GS6: I'm looking at the time, the number of cells, three test tubes, temperature needs to go there, and so it's at 22 degrees and 10 degrees Celsius, and as the time progresses we will see whether there is any growth of bacteria or not so at 22 degrees I see there is growth and at 10 degrees there is not as much. UGR1: Okay so measurements of the number of leaves are taken every thirty hours for up to 120 hours. Looks like they have three plants in each treatment group. UGNR1: So we are doing this at 22 degrees Celsius and 10 degrees Celsius.

GS, graduate student; P, professor; UGNR, undergraduate student who did not have research experience; UGR, undergraduate student who did have research experience.

professors and UGRs. The theme graph choice was observed for multiple GSs and UGRs, but only for one UGNR. Professors were unique in that they were the only group who did not explicitly state the graph choice in the planning phase.

Finally, within the MRC category learning/reflection, the theme data table appeared with multiple subjects and across all participant groups.

Construction Phase

The construction phase followed the planning phase and began with the drawing of the axes and ended when a participant signaled that he or she had finished constructing the graph. Relative to the planning phase, most participants spent more time constructing their graphs (Figure 1). However, professors spent less time than the other three participant groups. This is consistent with the graphs they created (see *Graph Attributes*). Although each participant constructed a graph, some of the par-

ticipants regurgitated the information presented in the data table and focused on plotting points, labeling axes, titling the graph, making a key, and scaling the axes.

All four MRC categories were present in the construction phase, with a focus on invention (Figure 2). Ideally, as participants were constructing their graphs, they also should have been reflecting on their graph choice, critiquing the data provided, and ending with a take-home message of the data they just plotted. A summary of the MRC categories, themes, and examples from transcripts is displayed in Table 4. Compared with the planning phase, there was more diversity in the distribution of themes across the MRC categories and across the participant groups during the construction phase (Figure 2).

Themes within the MRC category invention were similar across the participant groups and were data type, statistics, and graph construction. However, the theme statistics was seen only for one UGR. Within the MRC category critique, sample

TABLE 4. Construction phase: summary of the themes, definitions, and participant examples

Categories in MRC	Themes	Participant examples
Function	Graph choice: participant is explicitly stating graph choice (i.e., bar, line, scatter) based on the data provided in the table. Participants may also interject their personal feelings or rely on their past experiences when contemplating between different graph types, their usage, and limitations.	GS1: Oh that's a good point, whether or not I can connect them, because [with] the [variable] time line can be discrete. I'm not sure. I think since its cell growth over time that should be fine [to do] so (connects points on the graph). UGR2: I'm using the line graph because it shows the trend the easiest, because it goes straight and up a little. UGR3: … did I say line or bar? I'm doing lines. I'm doing a line chart now I changed my mind.
Invention	Statistics: participant is talking about either descriptive or inferential statistics.	P5: [the trend] is almost linear and [this is] because there is some error [in the data] which I didn't calculate (sketches error bars on each data point). P2: You do need a bigger sample size, but [I will estimate] the error [bar] for each one [treatment]. (adds error bars and labels lines as either 10C or 22C). GS7: I think what I'm going to do is take average of three tubes and make a bar for each time point at each temperature. I'm plotting to show the standard deviation from the average value. UGR4: … you can create a trendline for each dataset, so basically out of 15 ml and 5 ml, you can do the line of best fit, where you try and roughly go through as many of the points as possible. UGNR7: This graph looks like it's not going to be linear, but I'll make a line of best fit for each [tube] just so you can tell where it's going.
	Data type: participant is explicitly making decisions about whether or not to plot raw data or plot manipulated data (i.e., average) and the number of graphs to use to properly convey the data.	P1: I'm collapsing across tubes, so I'm giving total [number of cells], or I could do mean [number of cells]. GS5: There are three tubes within each temperature group, so I will do the average—calculate the mean of the number of cells for the same time point for all three tubes. And for each time point I can have the mean and standard deviation. UGR3: Okay well I'm going to make two charts then if that's the case. I'll make one the cell count at 22 degrees Celsius, and I'll make another one for cell count at 10 degrees Celsius with the same axes. UGR4: Because we have three plants, which is like three trials for each, I'm going to average the number of leaves at each time for each plant for each amount of water. UGNR6: I'm thinking maybe I could do like an average number of plants that would require doing calculations. There's fifteen milliliters of water a day. I'm just going to go ahead and do averages.
Learning/ reflection	Evaluation: participant is talking either about the general graphing habits, future directions, or take-home message.	P2: You do need a bigger sample size, but [I will estimate] the error [bar] for each one [treatment] (adds error bars and labels lines as either 10C or 22C). GS8: This is the most horrible graph ever because it's not even clear what the data mean. It might be easy for me to understand what I've done but it's not easy. If I gave it to you, I'm sure you would not understand it, if it was out of context. UGR4: You can see really clearly that they [lines] are increasing at the same rate but throughout the entire experiment, the 5 ml produces less leaves. UGNR3: I did this wrong … I should have put ml on the y axis … I'll just keep going with this. I might be okay … okay yeah I need to plot this with number of leaves instead of ml [scratches the x-axis label and renames it number of leaves]. The number of leaves will be on the x axis.
	Technology: participant is mentioning the habitual use of graph-making software to reflect on elements of the current graph construction.	GS3: So if I read the problem and use Excel, I can just put linear regression lines and the r^2 values, both are greater than 0.8 or something (draws 2 linear regression lines through points and labels lines with $r^2 > 0.8$). UGR1: So I feel like if I was doing this in Excel, I would make each plant its own representation symbol or its own color to better represent that. Have like a uniform structure to this but a different representation. UGR4: … if you are in Excel, you can [get] the equation for the trend line and it will tell you that y equals some function of x. From that, you can see the mathematical relationship behind the number of leaves that you have.

TABLE 4. Continued

Categories in MRC	Themes	Participant examples
Critique	Aesthetics: participant is using elements of graph design (i.e., gestalt principles and color) to critique the constructed graph.	UGR2: I guess I will graph the other [tubes] too and we can just imagine that they are different colors. UGNR1: I'd use different colors for the ones at 22 [degrees Celsius] and the ones at 10 [degrees Celsius] and then you can show that in the legend …. But the legend is black so I guess I'll just graph the points at different lines. They will all be the same color.
	Sample size: participant is critiquing the small sample size presented in the data table.	P4: With 3 plants in each, I guess you could put a standard error on that [data point]; $n = 3$ is pretty small but sometimes in biology, you are stuck with pretty small. I can't [calculate standard error] in my head but, what I would probably do is put each standard error at each [data] point, plus or minus. P2: You do need a bigger sample size. UGNR6: I'll draw the dotted line that represents the five milliliters of water per day, which is also approximately a linear line but if there was more data it could possibly be curving off to give a constant average, at least if you want any of those.

GS, graduate student; P, professor; UGNR, undergraduate student who did not have research experience; UGR, undergraduate student who did have research experience.

size was seen for multiple professors, but only one UGNR. Professors critiqued the data presented and indicated that a bigger sample size would be preferable to run inferential statistics (Figure 2). However one professor connected the small sample size to a possible real-life situation a biologist could encounter, saying "With 3 plants in each, I guess you could put a standard error on that, $n = 3$ is pretty small but sometimes in biology, you are stuck with pretty small." The theme aesthetics emerged for multiple UGNRs and one UGR, but did not appear for professors or GSs. Within the MRC category function, only one theme, graph choice, emerged for GSs, UGRs, and UGNRs. Within the MRC category of learning/reflection, the themes technology and evaluation emerged. While evaluation was prevalent for multiple participants across all groups, technology was only present in GSs and UGRs.

Reflection Phase, Graph Choice
The reflection phase followed the construction phase and began when the interviewer intervened and probed the participants to elaborate on their graph choice and what they plotted. Figure 1 displays the amount of time the participants spent answering the reflection question "Why did you choose to make this type of graph?" There was a significant difference in the amount of time spent reflecting between GSs and UGRs ($p < 0.05$; independent-samples t test, SPSS v. 22) and GSs and UGNRs ($p < 0.01$; independent-samples t test, SPSS v. 22).

All four MRC categories were present in the reflection phase, which specifically targeted the learning and reflection category. We expected participants to elaborate on graph choice, using the graph created in the construction phase (invention) to provide a reflection and critique. A summary of the MRC categories, themes, and examples from transcripts are displayed in Table 5.

All participants provided an answer for this phase, and the most prevalent theme across the participant groups was evaluation, which is not surprising, because the participants were probed to reflect on their graph choice. However, there were different reasoning categories under this theme. Four UGNRs and four GSs used their personal experiences and intuitions when reflecting on their graph choice; two UGNRs, two GSs, and professors used this opportunity to justify their graph choice by explaining why bar, pie, and scatter plots would not accurately display the data; two UGNRs and one GS formulated the take-home message for the graph; and the other two used the data table to justify their reasoning for constructing a line graph—a theme that was not seen in the professor group and was only seen with one UGR and GS. It is also interesting to note that the themes purpose and variables were present only in the GS and professor populations. The professors stated the purpose of the experiment and aligned it with the message portrayed by their graph.

All of the participants who mentioned time in their reflection constructed line graphs. We did not notice differences in the participants' graph reflection themes and the graphing scenarios.

Overall Patterns In Graph-Construction Reasoning
The distribution of themes within the MRC categories and across the participant groups was the most diverse in the construction and reflection phases (Figure 2). Across all the MRC categories, there were multiple instances in all population groups when all three themes under the MRC category of invention were mentioned in either the planning, construction, or reflection phases (see Figure 2). In the MRC category function, the theme graph choice appeared for all participant groups and multiple times either in the planning, construction, or reflection phases. Another theme that was well represented across the construction and reflection phases for all participant groups was evaluation, and it fell under the MRC learning and reflection category. A second theme under this same category, data table, was common across all participant groups, but only in the planning phase. Remaining themes under the MRC categories critique, function, and learning and reflection were less frequent.

GRAPH ATTRIBUTES
To address our second research question aimed at characterizing the quality and attributes of graphs constructed by participants, we described the graphs qualitatively based on similarities and differences that emerged across participants and participant groups (Table 1, Figure 3, and Figures 2–8 in the Supplemental Material).

Graphs constructed by undergraduate students (UGRs and UGNRs) and graduate students (GSs), but not professors,

TABLE 5. Reflection phase: summary of the themes, definitions, and participant examples

Categories in MRC	Themes	Participant examples
Function	Purpose: this is when the participant explicitly states that the purpose of the graph is to align with the purpose of the task.	P2: Well you want to see the effect of temperature on growth. Here (pointing to the graph), you can easily see the two treatments, [and the] two levels of temperatures that were used while they changed over time. GS4: My question was how temperature affects the growth of bacteria, so here I can see the difference between these two lines is how much difference the temperature had on the growth.
	Time: participant is using phrases like "change over time" or "flow over time" to justify choosing a line graph.	P5: I would say that [usually] [when you have the variable] time, a line graph is used. GS1: I would be able to show how the cell number changed over time. UGR1: Things that are measuring changes over time I think lines show trends there better than my initial thought of a bar graph.
	Variables: participant explains variables in the data table using the words "independent" or "dependent."	P5: So because we have independent variable, time and dependent variable, number of leaves and we have two—in this case, two different conditions of, uh, amount of water that a second variable and we can just show it as two different lines. GS1: I was trying to decide whether or not time was going to be a continuous variable. I ended up thinking it would be, even though it might not be because of the distinct chosen time points.
Invention	Statistics: participant is talking about either descriptive or inferential statistics.	P2: Of course we know that as more time passes bacteria grow faster, but there could be an interaction between time and temperature [not depicted by the data plotted]. GS3: … in the beginning I was thinking [of] putting the standard deviations but I decided to [plot] the data first [and] I think that putting a linear regression is very easy to use and read. UGR4: You can't compare the number of leaves for 15 ml at 120 hours with the 5 ml at 30 hours because that's just not a fair comparison. You have to show them linearly and in some kind of relationship. UGNR6: A best fit line is like when you have points that almost make a linear line but they're a little bit off which could be due to experimental error. So you draw a line that best represents all the data so it doesn't go minimum and a maximum so it kind of evens it out if you have some equal number of points below the best fit line and above, so it makes an average between the line.
Learning/ reflection	Evaluation: participant is talking about the general graphing habits, future directions, or take-home message.	P5: (Pointing to the graph) If this [was] 4 different plants instead of time points then I probably would have [made] a bar graph, [to accommodate for] more categories. GS8: If I were to do any other type of bar graph or something, I'm not very sure how to do that by myself. Maybe if I were to do it in Excel then, yeah. The truth is, I don't really know what type of data to use for a bar graph. UGR4: One of the scales in the experiment was the passing of time. You can't use a bar graph or pie chart to show the passing of time, because you're going to want to show it like linearly along some kind of axis, so that means you're going to have to find some way to put the data points sequentially according to the time it happened, in order to compare them accurately. UGNR1: This is the most common type of graph that I make so I thought of this kind first.
	Data table: this is when the participant is making sense of the data provided in the data table as evidenced by summarizing the data and or the variables presented.	GS1: … since the two variables have the same cell number over time, things that are being studied could both be displayed on the same graph which would help to visualize by looking at one time point, [which is] why I chose the line [graph]. UGR1: The way this chart is presented, at first I thought it was a comparison because plant 1,2, and 3 is redundant, but that's just in my treatment group so I misread that. UGNR3: Because in order to plot time versus number of leaves, you'd have to do a scatter plot of sorts. In retrospect, I should have made two graphs and separated them out into 5 and 15 ml. UGNR1: Because that's what I thought about when I first looked at this chart and it does show the number of cells.
Critique	Aesthetics: participant is using elements of graph design (e.g., gestalt principles and color) to critique the constructed graph.	GS8: I know that if I were to make this graph in Excel, I could put in a lot of colors and make sense out of it. UGR1: Ideally, this would be a little bit more visually appealing with different colors and evenly spaced dots and lines.

GS, graduate student; P, professor; UGNR, undergraduate student who did not have research experience; UGR, undergraduate student who did have research experience.
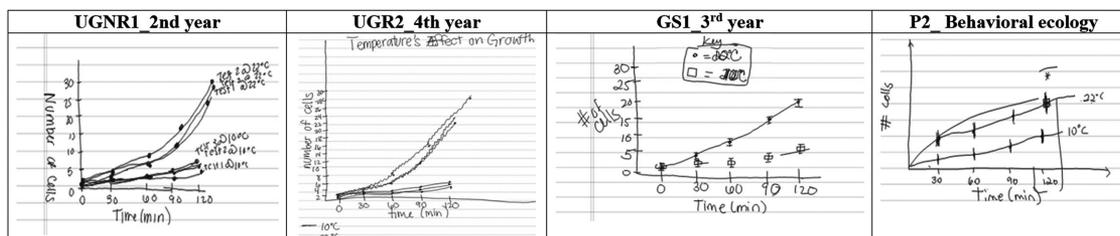
FIGURE 3. Graph exemplars from all participant groups using the bacteria scenario. (A complete summary of the graphs constructed can be found in Figures 1–8 of the Supplemental Material.)

followed basic graph conventions and included meticulously labeled axes, titles, tick marks, scale, and key. Ten of the 15 undergraduate students titled their graphs, whereas only one of the eight GSs and one of the five professors titled their graphs. In terms of axis labels, all participants labeled their axes appropriately based on the data they chose to plot with time on the x-axis and either number of leaves or cells on the y-axis. However, one UGNR struggled with labeling the axes, initially having a difficulty deciding how to organize the axes and label them such that the independent variable, time, is on the y-axis instead of the x-axis. Almost all participants indicated time in either minutes or hours. All participants had an appropriate scale, except for Professor 2, who did not scale the y-axis. Two students did not plan ahead concerning the space they needed for the scale, realizing midway through the scaling process that they were running out of space, so they decided to add axis breaks (Figures 1–8 of the Supplemental Material). In contrast to the undergraduate and graduate students, professors tended to sketch their graphs, omitting detailed axis labels and meticulous plotting (Figures 1–8 of the Supplemental Material).

Of the 15 undergraduate students, eight plotted all of the raw data points, four plotted some of the raw data, and three plotted averages. In contrast, graduate students and professors and three undergraduate students collapsed the data, plotted transformed data values, and sketched error bars (descriptive statistics) or mentioned a statistical test they would run (inferential statistics) to show meaningful trends and changes.

Participants who were randomly assigned the bacteria scenario generally constructed a line graph, except for three graduate students who constructed either a scatter or a bar graph (Figures 1–8 of the Supplemental Material). Line graphs represent the general consensus for this scenario in biology textbooks (e.g., Freeman *et al.*, 2017) and primary literature (e.g., Ratnowsky *et al.*, 1982; Zwietering *et al.*, 1990, 1991), because they are associated with either logistic or exponential growth models. There are also studies that report data on bacterial growth with temperature in bar graphs, box-and-whisker plots, and categorical dot plots (e.g., Seel *et al.*, 2016). There was greater variety among the graphs constructed by participants who were randomly assigned the plant scenario (Figures 2, 4, 6, and 8 of the Supplemental Material). These results are similar to the bar and line graphs displayed by Mayak *et al.* (2004), looking at how water affects plant growth. In our study, we did not see specific themes that were exclusive to either only the bacteria or the plant scenario. We

did notice that some of the participants who constructed a line graph used the theme time in their graph reflection.

The graphs constructed by all participants were, in general, aesthetically sound, and the presence of gestalt principles (i.e., proximity, continuity, and connectedness) enabled easy observation of the general data trends and take-home message. The ink-to-white space was appropriate, and what was plotted was clear without extraneous elements. However, there were five graphs that had too many lines with overlapping data point labels, which made it difficult to understand the take-home message. In particular, the graph constructed by UGNR3 was sufficiently unclear that the viewer found it difficult to identify the data points and formulate a clear take-home message (Figures 1–4 of the Supplemental Material).

An important purpose of graphs that summarize data is the alignment of the data presented and graph chosen with the research question and/or hypothesis. In our interview task, this was looking at either how temperature affects the growth of bacteria or how the amount of water influences plant growth. The graphs of four undergraduate students did not align with the research question or hypothesis, as only a subset of the data was plotted (e.g., data from one treatment). All graphs constructed by graduate students aligned with the research question posed in the task.

DISCUSSION

In this study, we used the MRC framework to understand how undergraduate students, graduate students, and professors reason with graph choice, data, and graph construction and how the attributes of the graphs constructed by the study participants might differ.

Implicit in the MRC framework is expert competence with creating and understanding external representations. While all participants engaged in reasoning within all MRC categories, there is evidence for expert–novice differences across our participant groups (Figure 2). All professors took time to understand the data before proceeding with graph construction, and all but one graduate student planned, whereas only some of the undergraduate students planned before proceeding with graph construction. Generally, we saw that, when reflecting on their graphs, expert professors focused on the function of the graph and showcased their understanding with concepts related to experimental design, while novice undergraduate students generally relied on their intuition and data given to them in the task. We also saw expert–novice differences in the data plotted in the graphs of undergraduate

Novice → Expert

- Minimal to no planning prior to construction
- Meticulous graph construction
- Raw data often used in graphs
- Data considered independent of experimental context
- Variability in experimental data not considered
- Graph choice not based on question and/or hypothesis
- Intuitive reasoning about data variables and experimental replicates
- Reflection focused on superficial graph features and aesthetics

- Extensive planning with decisions on data type and purpose prior to construction
- Data transformed before graphing
- Less time on construction and data plotting aimed to note trends
- Data considered within experimental context (e.g. sample size, replicates, variables)
- Variability in experimental data important
- Graph choice based on question and/or hypothesis
- Experience and disciplinary expertise-based choices
- Reflection focused on experimental concepts, statistics, and graphing purpose
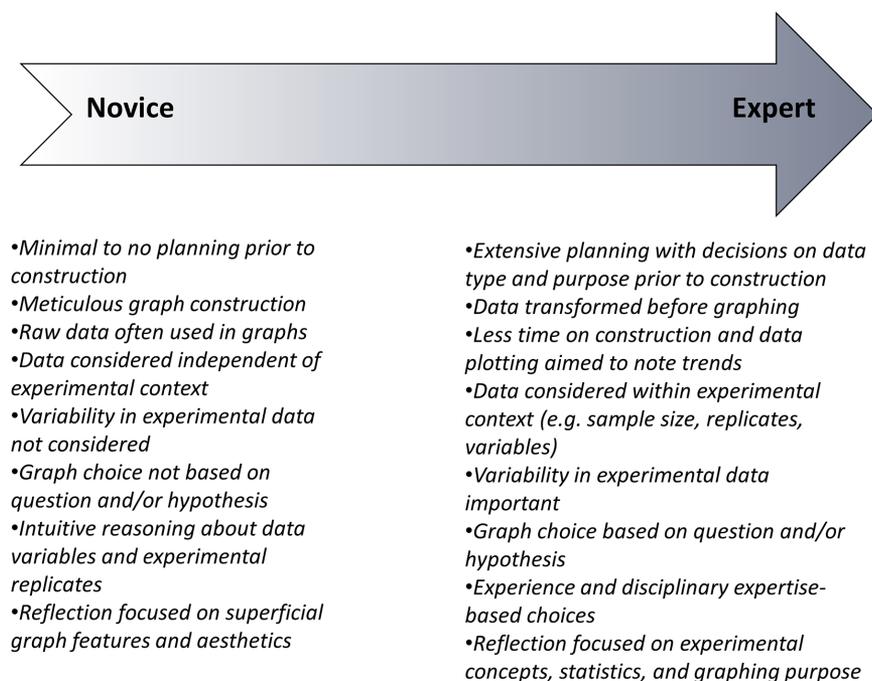
FIGURE 4. Visual summary of graph-construction reasoning, graphing behavior, and graph attribute findings with the reasoning behind graph choice and construction, graphing behaviors, and graph attributes along the novice to expert continuum.

students, graduate students, and professors. Most undergraduate students meticulously plotted all raw data, whereas most professors and graduate students plotted transformed data values. Our data are reminiscent of an expert–novice study conducted in the context of neurobiology that also noted differences in drawing of neurons by undergraduate students, graduate students, and laboratory leaders (professors; Hay *et al.*, 2013). Undergraduate students' representations were meticulous reproductions of neurons illustrated in textbooks. Neuron drawings by graduate and postdoctoral students closely resembled images seen under the microscope and were influenced by observations from their research projects, whereas the expert laboratory leaders used years of research experience to create imaginative drawings based on hidden hypotheses. Findings reported by Hay *et al.* (2013) and our graphing study are supported by the National Research Council (2000), which states that experts organize their knowledge in a way that reflects a deep understanding of the subject matter and expert knowledge cannot be recalled as a set of isolated facts but is applied to the context or the problem that is being solved. Deep understanding is evident in professors' graph reflections as they talk about the purpose of the graph, experimental design, and relevant concepts that are not present in the reasoning of the undergraduate students. Jordan *et al.* (2011) found that, when solving a task, experts were more likely to use their prior knowledge and discuss ideas at a broader context as compared with novices, who solved the task with only the information given to them. Likewise, in the Hay *et al.* (2013) study, neuron drawings by the laboratory leaders were original and unlike those found in textbooks, because the experts' drawings were informed by years of experience and accumulated knowledge.

IMPLICATIONS FOR INSTRUCTORS

Our study revealed that, while all participant groups showed evidence of reasoning within all MRC categories, the identity of that reasoning was often different in a manner that is consistent with expected expert–novice differences as highlighted earlier. Further, the graphs produced by participants in the study also varied along the novice–expert continuum. Figure 4 summarizes the graph-construction reasoning, behaviors, and graphs that we observed in the most novice and most expert participants. The distinctions summarized in this figure highlight the beginning of hypothetical learning trajectories and potential target areas for instructors to promote more expert-like reflective data handling and graphing practices. As more undergraduate students are encouraged to engage in inquiry and research project–based biology labs and seek research apprenticeship opportunities during their higher education, they will be engaged in the scientific practice of data analysis and presentation. Therefore, it is important to provide students with targeted instruction that not only advances their biology content knowledge but also facilitates their data handling and representation skills toward expertise. While students have experience with graphing dating back to elementary school, our data suggest that refocusing and scaffolding their data handling and graphing activities in the context of their undergraduate learning experiences is needed. Kim and Hannafin (2011) suggest designing and implementing instructional scaffolds that target student difficulties with conceptual, procedural, metacognitive, and strategic knowledge (Kim and Hannafin, 2011).

Conceptual scaffolds, as they relate to graphing, can structure students' understanding of the purpose of a graph and allow them to gauge their graph knowledge. Sketching a graph to visualize concepts in experimental design is an approach suggested by Dasgupta *et al.* (2014). Procedural scaffolds help students learn the stepwise procedures that underlie graph choice and construction. There are many published examples that emphasize taking a procedural approach to graphing (Kosslyn, 1994; Paniello *et al.*, 2011; Webber *et al.*, 2014; Duke *et al.*, 2015). Metacognitive scaffolds allows students to monitor their problem-solving processes with a focus on constant reflection (Kim and Hannafin, 2011). We published a tool (see Step-by-Step Guide in Angra and Gardner, 2016) that helps students plan their data, construct graphs, and then reflect on their graphs in a methodical manner. This tool is a metacognitive scaffold (Kim and Hannafin, 2011), because it contains the reflection piece after graph construction. Even in this study, the interviewer followed up with participants with reflective questions asking about graph choice. In a classroom setting, instructors can include reflective prompts throughout multiple assignments to help students develop their metacognitive abilities. The last scaffolding strategy, strategic scaffolds, challenges

students to consider other options as they are solving problems. Although previously published graphing materials provide students with many examples of graphs, these resources do not provide explicit strategic scaffolding, because they do not ask students to consider other options.

Using these tools and scaffolding strategies to emphasize graph choice and construction skills will encourage students to think critically about data and graphs in and outside the classroom. This is important, because students are rarely asked to reflect critically on the affordances and limitations of representations that they choose (diSessa and Sherin, 2000). Incorporating these and other graphing materials during teacher education may provide teachers with tools to guide students successfully and confidently toward proper graph construction. This would be useful in undergraduate curricula as well, as has been suggested by a continuing education approach for biologists teaching statistical concepts (Weissgerber *et al.*, 2016).

PROJECT SCOPE AND FUTURE STUDIES

Four main study design features bounded the scope of our conclusions. First, data were collected from students and professors at a single midwestern U.S. research university, which is a unique environment with its own curriculum and student population. Furthermore, our study consisted of a small group of participants, so the claims we present are not broad generalizations to the types of things that all professors or students do or think. However, many of our findings are consistent with and extend from previous work by others. To verify our findings fully, future work is needed at other types of institutions, in different disciplinary fields, and with their own unique participants to fully understand and appreciate the reasoning behind graph choice and construction.

Second, we provided all participants with a simple data set with one independent variable, one dependent variable, and two treatments with three replicates each. For our study to be replicated in a different disciplinary context, the bacteria and plant scenarios would need to be modified to fit the appropriate purpose, with data and experimental methods that conform to the disciplinary norms and practices. However, the simple data set did confirm some previous difficulties documented in the literature. UGNR4 and UGR3 showed difficulty with scaling axes (Figures 1–8 of the Supplemental Material; Padilla *et al.*, 1986; Li and Shen, 1992; Brasell and Rowe, 1993; Ainley, 2000), as indicated by the awkward positioning of the axis breaks, and UGNR3 showed difficulty with variables, as indicated by the graph produced (Tairab and Al-Naqbi, 2004; Figures 1–8 of the Supplemental Material). However, the simplicity of the data set may have caused Professor 2 to go into "teacher mode" and quickly sketch the data to illustrate how temperature influences bacteria growth, instead of taking time to plot data.

Third, participants in our study were given a data set. Previous studies have shown that, when students use their own data to perform advanced tasks, they show deeper reasoning than when they use someone else's data (Kanari and Millar, 2004). A future study can examine graph choice and construction with a more elaborate data set and with data the participants collected themselves in CUREs or inquiry lab classes or with data from simulations.

Finally, participants in this study constructed graphs manually using a LiveScribe pen and paper instead of the modern and conventional method of graph construction on the computer. Having participants narrate their thought processes during manual construction allowed us to fully understand their reasoning. If we had asked participants to construct graphs using software programs, that request might have tampered with their graph choice by biasing them toward graph choices presented by the software package. Manual construction allowed us to slow participants down and probe their graph-construction reasoning fully. We do acknowledge that biologists at all levels of expertise rarely construct graphs for formal presentation by hand. However, informal communication with peers during instruction often involves the generation of quick, sometimes simplified graphs (Roth and Bowen, 2003). We saw evidence of this with our professor population, one professor in particular studied the data table and then sketched the data with error bars to answer the research question quickly. With the data from our simple task, we can now move to more complex data sets and digital environments to further reveal areas of difficulties and competencies with graphing.

REFERENCES

Ainley, J. (2000). Transparency in graphs and graphing tasks: An iterative design process. *Journal of Mathematical Behavior, 19,* 365–384.

Ainley, J., Nardi, E., & Pratt, D. (2000). The construction of meanings for trend in active graphing. *International Journal of Computers for Mathematical Learning, 5,* 85–114.

Ali, N., & Peebles, D. (2011). The different effects of thinking aloud and writing on graph comprehension. In *Proceedings of the twentieth annual conference of the Cognitive Science Society* (pp. 3143–3148). Mahwah, NJ: Erlbaum.

Aliaga, M., Cobb, G., Cuff, C., Garfield, J., Gould, R., Lock, R., & Witmer, J. (2005). Guidelines for Assessment and Instruction in Statistics Education: College Report. Retrieved August 9, 2016, from www.amstat.org/asa/files/pdfs/GAISE/2005GaiseCollege_Full.pdf

American Association for the Advancement of Science. (2011). *Vision and change in undergraduate biology education: A call to action.* Washington, DC.

Angra, A., & Gardner, S. M. (2016). The development of a framework for graph choice and construction. *Advances in Physiology Education, 40,* 128–128.

Association of American Medical Colleges. (2009). *Scientific foundations for future physicians.* Washington, DC.

Auchincloss, L. C., Laursen, S. L., Branchaw, J. L., Eagan, K., Graham, M., Hanauer, D. I., ... Dolan, E. L. (2014). Assessment of course-based undergraduate research experiences: A meeting report. *CBE—Life Sciences Education, 13,* 29–40.

Bangera, G., & Brownell, S. E. (2014). Course-based undergraduate research experiences can make scientific research more inclusive. *CBE—Life Sciences Education*, *13*, 602–606.

Beichner, R. J. (1994). Testing student interpretation of kinematics graphs. *American Journal of Physics*, *62*, 750–762.

Berg, C. A., & Smith, P. (1994). Assessing students' abilities to construct and interpret line graphs: Disparities between multiple-choice and free-response instruments. *Science Education*, *78*, 527–554.

Bowen, G. M., & Roth, W. M. (2005). Data and graph interpretation practices among preservice science teachers. *Journal of Research in Science Teaching*, *42*, 1063–1088.

Brasell, H. M., & Rowe, M. B. (1993). Graphing skills among high school physics students. *School Science and Mathematics*, *93*, 63–70.

Bray-Speth, E., Momsen, J. L., Moyerbrailean, G. A., Ebert-May, D., Long, T. M., Wyse, S., & Linton, D. (2010). 1, 2, 3, 4: Infusing quantitative literacy into introductory biology. *CBE—Life Sciences Education*, *9*, 323–332.

Bright, G. W., & Friel, S. N. (1998). Graphical representations: Helping students interpret data. In Lajoie, S. P. (Ed.), *Reflections on statistics: Learning, teaching, and assessment in grades K–12* (pp. 63–88). Mahwah, NJ: Lawrence Erlbaum Associates.

Brownell, S. E., Hekmat-Scafe, D. S., Singla, V., Seawell, P. C., Imam, J. F. C., Eddy, S. L., ... Cyert, M. S. (2015). A high-enrollment course-based undergraduate research experience improves student conceptions of scientific thinking and ability to interpret data. *CBE—Life Sciences Education*, *14*, ar21.

Bruno, A., & Espinel, M. C. (2009). Construction and evaluation of histograms in teacher training. *International Journal of Mathematical Education in Science and Technology*, *40*, 473–493.

Chi, M. T. H. (2006). Two approaches to the study of experts' characteristics. In Ericsson, K. A., Charness, N., Feltovich, P. J., & Hoffman, R. R. (Eds.), *Cambridge handbook of expertise and expert performance* (pp. 21–30). Cambridge, UK: Cambridge University Press.

Clase, K. L., Gundlach, E., & Pelaez, N. J. (2010). Calibrated peer review for computer-assisted learning of biological research competencies. *Biochemistry and Molecular Biology Education*, *38*, 290–295.

Cleveland, W. S. (1984). Graphical methods for data presentation: Full-scale breaks, dot charts, and multibased logging. *American Statistician*, *38*, 270–280.

Colon-Berlingeri, M., & Burrowes, P. A. (2011). Teaching biology through statistics: Application of statistical methods in genetics and zoology courses. *CBE—Life Sciences Education*, *10*, 259–267.

Cooper, R. J., Schriger, D. L., & Close, R. J. (2002). Graphical literacy: The quality of graphs in a large-circulation journal. *Annals of Emergency Medicine*, *40*, 317–322.

Cooper, R. J., Schriger, D. L., & Tashman, D. A. (2001). An evaluation of the graphical literacy of Annals of Emergency Medicine. *Annals of Emergency Medicine*, *37*, 13–19.

Creswell, J. W. (2013). *Qualitative inquiry and research design: Choosing among five approaches*. Thousand Oaks, CA: Sage.

Dasgupta, A. P., Anderson, T. R., & Pelaez, N. (2014). Development and validation of a rubric for diagnosing students' experimental design knowledge and difficulties. *CBE—Life Sciences Education*, *13*, 265–284.

diSessa, A. A. (2004). Metarepresentation: Native competence and targets for instruction. *Cognition and Instruction*, *22*, 293–331.

diSessa, A. A., & Sherin, B. L. (2000). Meta-representation: An introduction. *Journal of Mathematical Behavior*, *19*, 385–398.

Drost, E. A. (2011). Validity and reliability in social science research. *Education Research and Perspectives*, *38*, 105.

Duke, S. P., Bancken, F., Crowe, B., Soukup, M., Botsis, T., & Forshee, R. (2015). Seeing is believing: Good graphic design principles for medical research. *Statistics in Medicine*, *34*, 3040–3059.

Ericsson, K. A. (2006). The influence of experience and deliberate practice on the development of superior expert performance. In Ericsson, K. A., Charness, N., Feltovich, P. J., & Hoffman, R. R. (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 683–703). New York: Cambridge University Press.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis*. Cambridge, MA: MIT Press.

Few, S. (2004). *Show me the numbers: Designing tables and graphs to enlighten*. Oakland, CA: Analytics Press.

Franzblau, L. E., & Chung, K. C. (2012). Graphs, tables, and figures in scientific publications: The good, the bad, and how not to be the latter. *Journal of Hand Surgery*, *37*, 591–596.

Freeman, S., Quillin, K., & Allison, L. (2017). *Biological science* (6th ed.). Glenview, IL: Pearson Higher Ed, Pearson Education.

Friel, S. N., & Bright, G. W. (1996). Building a theory of graphicacy: How do students read graphs? Paper presented at the Annual Meeting of the American Educational Research Association, held April 8–12, 1996, New York (ERIC Doc. Reproduction Service No. ED 395277).

Grawemeyer, B., & Cox, R. (2004). The effect of knowledge-of-external-representations upon performance and representational choice in a database query task. In Blackwell, A., Marriott, K., & Shimojima, A. (Eds.), *Diagrammatic representation and inference, Third International Conference, Diagrams 2004* (pp. 351–354).

Grunwald, S., & Hartman, A. (2010). A case-based approach improves science students' experimental variable identification skills. *Journal of College Science Teaching*, *39*, 28.

Harsh, J. A., & Schmitt-Harsh, M. (2016). Instructional strategies to develop graphing skills in the college science classroom. *American Biology Teacher*, *78*, 49–56.

Hatch, J. A. (2002). *Doing qualitative research in education settings*. Albany, NY: SUNY Press.

Hay, D. B., Williams, D., Stahl, D., & Wingate, R. J. (2013). Using drawings of the brain cell to exhibit expertise in neuroscience: Exploring the boundaries of experimental culture. *Science Education*, *97*, 468–491.

Humphrey, P. B., Taylor, S., & Mittag, K. C. (2014). Developing consistency in the terminology and display of bar graphs and histograms. *Teaching Statistics*, *36*, 70–75.

IBM (2013). *SPSS statistics for Windows, version 22.0*, Armonk, NY.

Jordan, R. C., Ruibal-Villasenor, M., Hmelo-Silver, C. E., & Etkina, E. (2011). Laboratory materials: Affordances or constraints? *Journal of Research in Science Teaching*, *48*, 1010–1025.

Kanari, Z., & Millar, R. (2004). Reasoning from data: How students collect and interpret data in science investigations. *Journal of Research in Science Teaching*, *41*, 748–769.

Kellman, P. J. (2000). An update on gestalt psychology. In Landau, B., Sabini, J., Jonides, J., & Newport, E. (Eds.), *Perception, cognition, and language: Essays in honor of Henry and Lila Gleitman*. Cambridge, MA: MIT Press.

Kim, M. C., & Hannafin, M. J. (2011). Scaffolding problem solving in technology-enhanced learning environments (TELEs): Bridging research and theory with practice. *Computers & Education*, *56*, 403–417.

Konold, C., & Higgins, T. L. (2003). Reasoning about data. In Kilpatrick, J., Martin, G., & Schifter, D. (Eds.), *A research companion to principles and standards for school mathematics* (pp. 193–215). Reston, VA: NCTM.

Konold, C., Higgins, T., Russell, S. J., & Khalil, K. (2015). Data seen through different lenses. *Educational Studies in Mathematics*, *88*, 305–325.

Kosslyn, S. M. (1994). *Elements of graph design*. New York: Freeman.

Kostelnick, C. (1998). Conflicting standards for designing data displays: Following, flouting, and reconciling them. *Technical Communication*, *45*, 473–473.

Leinhardt, G., Zaslavsky, O., & Stein, M. K. (1990). Functions, graphs, and graphing: Tasks, learning, and teaching. *Review of Educational Research*, *60*, 1–64.

Leonard, J. G., & Patterson, T. F. (2004). Simple computer graphing assignment becomes a lesson in critical thinking. *NACTA Journal*, *48*, 17–21.

Li, K. Y., & Shen, S. M. (1992). Students' weaknesses in statistical projects. *Teaching Statistics*, *14*, 2–8.

Linn, M. C., Palmer, E., Baranger, A., Gerard, E., & Stone, E. (2015). Undergraduate research experiences: Impacts and opportunities. *Science*, *347*, 1261757.

LiveScribe (2015). LiveScribe Smartpens. Retrieved August 9, 2016, from www.livescribe.com/en-us/smartpen

Mayak, S., Tirosh, T., & Glick, B. R. (2004). Plant growth-promoting bacteria that confer resistance to water stress in tomatoes and peppers. *Plant Science*, *166*, 525–530.

McFarland, J. (2010). Teaching and assessing graphing using active learning. *MathAMATYC Educator*, *1*, 32–39.

Metz, A. M. (2008). Teaching statistics in biology: Using inquiry-based learning to strengthen understanding of statistical analysis in biology laboratory courses. *CBE—Life Sciences Education*, *7,* 317–326.

National Institutes of Health. (2016). Research Training and Career Development. Retrieved August 9, 2016, from https://researchtraining.nih.gov/programs/training-grants

National Research Council. (2000). *How people learn*. Washington, DC: National Academies Press.

National Science Foundation. (2016). Research Traineeship Program. Retrieved August 9, 2016, from www.nsf.gov/funding/pgm_summ.jsp?pims_id=505015

Padilla, M. J., McKenzie, D. L., & Shaw, E. L. (1986). An examination of the line graphing ability of students in grades seven through twelve. *School Science and Mathematics*, *86,* 20–26.

Paniello, R. C., Neely, J. G., Rich, J. T., Slattery, E. L., & Voelker, C. C. (2011). Practical guide to choosing an appropriate data display. *Otolaryngology—Head and Neck Surgery*, *145,* 886–894.

Patterson, T. F., & Leonard, J. G. (2005). Turning spreadsheets into graphs: An information technology lesson in whole brain thinking. *Journal of Computing in Higher Education*, *17,* 95–115.

Patton, M. Q. (2001). *Qualitative research and evaluation methods*. Thousand Oaks, CA: Sage.

Picone, C., Rhode, J., Hyatt, L., & Parshall, T. (2007). Assessing gains in undergraduate students' abilities to analyze graphical data. *Teaching Issues and Experiments in Ecology*, *5,* 1–54.

Polya, G. (1945). *How to solve it: A new aspect of mathematical method*. Princeton, NJ: Princeton University Press.

Preece, J., & Janvier, C. (1992). A study of the interpretation of trends in multiple curve graphs of ecological situations. *School Science and Mathematics*, *92,* 299–306.

Ratkowsky, D. A., Olley, J., McMeekin, T. A., & Ball, A. (1982). Relationship between temperature and growth rate of bacterial cultures. *Journal of Bacteriology*, *149,* 1–5.

Reeves, T. D., Marbach-Ad, G., Miller, K. R., Ridgway, J., Gardner, G. E., Schussler, E. E., & Wischusen, E. W. (2016). A conceptual framework for graduate teaching assistant professional development evaluation and research. *CBE—Life Sciences Education*, *15,* es2.

Roth, W. M., & Bowen, G. M. (2001). Professionals read graphs: A semiotic analysis. *Journal for Research in Mathematics Education*, *32,* 159–194.

Roth, W. M., & Bowen, G. M. (2003). When are graphs worth ten thousand words? An expert-expert study. *Cognition and Instruction*, *21,* 429–473.

Roth, W. M., & McGinn, M. K. (1997). Graphing: Cognitive ability or practice? *Science Education*, *81,* 91–106.

Rougier, N. P., Droettboom, M., & Bourne, P. E. (2014). Ten simple rules for better figures. *PLOS Computational Biology*, *10,* 1–7.

Schriger, D. L., & Cooper, R. J. (2001). Achieving graphical excellence: Suggestions and methods for creating high-quality visual displays of experimental data. *Annals of Emergency Medicine*, *37,* 75–87.

Schriger, D. L., Sinha, R., Schroter, S., Liu, P. Y., & Altman, D. G. (2006). From submission to publication: A retrospective review of the tables and figures in a cohort of randomized controlled trials submitted to the British Medical Journal. *Annals of Emergency Medicine*, *48,* 750–756.

Schussler, E., Torres, L. E., Rybczynski, S., Gerald, G. W., Monroe, E., Sarkar, P., … Osman, M. A. (2008). Transforming the teaching of science graduate students through reflection. *Journal of College Science Teaching*, *38,* 32–36.

Seel, W., Derichs, J., & Lipski, A. (2016). Increased biomass production by mesophilic food-associated bacteria through lowering the growth temperature from 30°C to 10°C. *Applied and Environmental Microbiology*, *82,* 3754–3764.

Seidman, I. (2013). *Interviewing as qualitative research: A guide for researchers in education and the social sciences*. New York: Teachers College Press.

Shah, P., Mayer, R. E., & Hegarty, M. (1999). Graphs as aids to knowledge construction: Signaling techniques for guiding the process of graph comprehension. *Journal of Educational Psychology*, *91,* 690–702.

Slutsky, D. J. (2014). The effective use of graphs. *Journal of Wrist Surgery*, *3,* 67–68.

Strauss, A., & Corbin, J. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory*. Los Angeles, CA: Sage.

Tairab, H. H., & Al-Naqbi, A. K. (2004). How do secondary school science students interpret and construct scientific graphs? *Journal of Biological Education*, *38,* 127–132.

Tanner, K. D. (2012). Promoting student metacognition. *CBE—Life Sciences Education*, *11,* 113–120.

Tufte, E. R. (1983). *Visual display of quantitative information*. Cheshire, CT: Graphic Press.

Wainer, H. (2013). *Medical illuminations: Using evidence, visualization and statistical thinking to improve healthcare*. Oxford, UK: Oxford University Press.

Webber, H., Nelson, S. J., Weatherbee, R., Zoellick, B., & Schauffler, M. (2014). The graph choice chart. *Science Teacher*, *81,* 37–43.

Weissgerber, T. L., Garovic, V. D., Milin-Lazovic, J. S., Winham, S. J., Obradovic, Z., Trzeciakowski, J. P., & Milic, N. M. (2016). Reinventing biostatistics education for basic scientists. *PLOS Biology*, *14,* 1–12.

Weissgerber, T. L., Milic, N. M., Winham, S. J., & Garovic, V. D. (2015). Beyond bar and line graphs: Time for a new data presentation paradigm. *PLOS Biology*, *13,* 1–10.

Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, *67,* 223–248.

Zwietering, M. H., De Koos, J. T., Hasenack, B. E., De Witt, J. C., & Van't Riet, K. (1991). Modeling of bacterial growth as a function of temperature. *Applied and Environmental Microbiology*, *57,* 1094–1101.

Zwietering, M. H., Jongenburger, I., Rombouts, F. M., & Van't Riet, K. (1990). Modeling of the bacterial growth curve. *Applied and Environmental Microbiology*, *56,* 1875–1881.