January 2016

Understanding, Evaluating, and Diagnosing Undergraduate Student Difficulties with Graph Choice and Construction

Aakanksha Angra
Purdue University

Graduate School Form
30 Updated 12/26/2015

PURDUE UNIVERSITY
GRADUATE SCHOOL
Thesis/Dissertation Acceptance

This is to certify that the thesis/dissertation prepared

By Aakanksha Angra

Entitled

UNDERSTANDING, EVALUATING, AND DIAGNOSING UNDERGRADUATE STUDENT DIFFICULTIES WITH GRAPH CHOICE AND CONSTRUCTION

For the degree of Doctor of Philosophy

Is approved by the final examining committee:

Stephanie M. Gardner
Chair

Nancy Pelaez

Edward Bartlett

Signe Kastberg

To the best of my knowledge and as understood by the student in the Thesis/Dissertation Agreement, Publication Delay, and Certification Disclaimer (Graduate School Form 32), this thesis/dissertation adheres to the provisions of Purdue University's "Policy of Integrity in Research" and the use of copyright material.

Approved by Major Professor(s): Stephanie M. Gardner

Approved by: Daoguo Zhou 9/22/2016

Head of the Departmental Graduate Program Date

UNDERSTANDING, EVALUATING, AND DIAGNOSING UNDERGRADUATE STUDENT DIFFICULTIES WITH GRAPH CHOICE AND CONSTRUCTION

A Dissertation

Submitted to the Faculty

of

Purdue University

by

Aakanksha Angra

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

December 2016

Purdue University

West Lafayette, Indiana

To my late grandmother and my parents for their unfailing love and endless support

ACKNOWLEDGEMENTS

I am truly grateful to my mentors, parents, and friends who have invested their time and efforts in my education. I will do my best to thank everyone I can in a few pages. I will start off by thanking my brilliant and supportive advisor, Dr. Stephanie Gardner, for her caring personality, infectious enthusiasm for teaching and passion for learning. I thank her for investing her time in me and training me in qualitative research, teaching me how to think and write, and allowing me to use her classroom to collect data for my dissertation work. I am truly fortunate to have her as a life-long friend and role model.

I thank my committee chair and head of Purdue International Biology Education Research Group (PIBERG), Dr. Nancy Pelaez, for inviting me to Purdue University to pursue a doctoral degree in biology education and encouraging me to pursue leadership roles in the biology department. I thank you for always supporting me in my research endeavors. Next, I thank Dr. Edward Bartlett for his persistent encouragement throughout my doctoral degree and discussions about potential career opportunities. I thank Dr. Signe Kastberg for her interest in my dissertation and for meeting with me to discuss the best ways to tell my research story to the public. I have truly appreciated her feedback and conversations about the graphing materials.

I thank my supportive co-instructor and friend, Dr. Elwood Walls, for his constant interest and help in collecting data and implement graphing materials in our lab sections over the years. I have really enjoyed working with him and will miss his stories! I would also like to thank my lab and office mates, Mozhu Li and Dr. Matthew Lira, for great conversations about the curveballs in life! I also thank past and current members of the PIBERG and Discipline Based Education Research for Graduate Students, Biology Graduate Student Council, and Graduate Research In Progress Series group for providing me with ample opportunities to present my research and giving me generous and honest feedback at the various stages of my research.

I want to thank all participants who took part in my research. You are all an invaluable part of my dissertation.

I also want to thank the staff and secretaries in the business, main office, and the graduate coordinator, Ms. Gina Rupp for making the graduate school process as smooth as possible by assisting with paperwork, teaching assistantships, and travel expense reports.

I also wish to acknowledge my mentors from my undergraduate institution who gave me opportunities to do research and teach from early on. Particularly, I thank my undergraduate advisor, Dr. Jeannette Yen for welcoming me in her marine biology lab and allowing me to work on various projects as an undergraduate student. I also thank Dr. Rachel Lasley-Rasher for being an excellent mentor, teaching me data curation skills, and for encouraging me to apply to graduate school in the first place! I thank my previous teaching mentor, Dr. Cara Gormally, who introduced me to the field of biology education research and told me about Purdue University's excellent biology education program. I would also like to thank the wonderful people I worked with at Georgia Tech with the PT-SURE program, who allowed me to achieve first-hand experience with designing curriculum for high school students, collecting data, and implementing curricula. I also thank Dr. Debi Huffman, who provided consistent guidance, encouragement, and introduced me to research and design in team settings.

Above all, I thank my parents for their love, sacrifice, and for constantly worrying about me. Thank you, dad, for having the biggest influence in my life, for emphasizing the importance of a graduate degree, and for celebrating every milestone with me. I thank my mom for always lending her ear and listening to me. I thank my sister for sending me hilarious memes and for always making me laugh. I thank my beautiful Sheltie, Rose, for her unconditional love, loyalty, and welcoming me during my visits home. I thank my best friend, Mike Melomed, for his love, devotion, encouragement, and patience during graduate school. Last but not the least, I thank my sweet, 3 year-old Bombay cat, Princess Luna Gump -- who always made sure I was punctual for all my meetings, willingly listened to my presentations, kept me company, and taught me the importance of napping. I could not have survived without you!

TABLE OF CONTENTS

Page

Page

Page

Page

ABSTRACT

Angra, Aakanksha. Ph.D., Purdue University, December 2016. Understanding, Evaluating, and Diagnosing Undergraduate Student Difficulties with Graph Choice and Construction. Major Professor: Stephanie M. Gardner.

Creating effective graphical representations of biological data is an essential component in the practices of science and involves engaging concepts and skills of quantitative literacy. With undergraduate biology students increasingly involved in scientific inquiry and experimentation, they are faced with the task of choosing and creating appropriate graphical representations of their data to communicate their findings. However, difficulties with graph choice and construction that were previously documented in literature, still exist today at both the K-12 and undergraduate levels. The purpose of this dissertation is to understand the reasoning involved behind choosing certain graph types and the process that occurs during graph construction, and to design and validate instructional materials to improve graphing skills. The first chapter reviews recent policy documents and relevant literature that have stressed the importance of graphing skill development. Although graphing has been heavily emphasized at the K-12 level and in the context of math and physics, the stepwise thought process and reasoning that determine how the graph is constructed and the final message it conveys are not well understood. In chapter two, I attempt to understand these reasoning that occurs during graph choice and construction by studying expert and novice biologists. Clinical think-

aloud interviews were conducted and participants were presented with a small data set and asked to construct a graph using pen and paper. In chapter three, I look at how graphs are constructed in a naturalistic, classroom setting. In Spring 2013 and 2014, students in an upper level physiology laboratory engaged in inquiry -based labs, which required them to work in a team to design experiments, collect data, and present these findings in an oral presentation. Students engaged in guided reflective practices multiple times over the course of the semester, which forced them to evaluate their graph choice and describe the advantages and the disadvantages of their graph. The work described in fourth chapter utilized findings from the second and third chapters, as well as existing literature to develop instructional and learning tools aimed at improving reasoning with graphs. These tools are: the step-by-step guide, guide to data displays, and the graph rubric. The step-by-step guide was informed by the data from the think-aloud interviews (chapter 2) and its purpose is to provide students with a framework for data presentation, as practiced by experts. The purpose of the guide to data displays is to inform students of various types of graphs, their usage, advantages, and disadvantages. The purpose of the graph rubric was to help instructors provide quick and consistent feedback on students' graphs and for students to use when constructing and critiquing graphs. The graph rubric was informed by: seminal literature in math and science education that informed the 12 assessment categories, expert-novice graphing interviews (chapter 2), and student graphs and reflections (chapter 3). The rubric was validated in three ways: assessing graphs from five introductory biology textbooks, graphs generated in the classroom, and graphs from the science literature. Chapter 5 used the cognitive apprenticeship model and tested the utility of the instructional and learning materials mentioned in chapter 4 in an upper-level physiology laboratory classroom (same setting and curriculum as chapter 3). Data for this chapter were collected during the Spring 2015 and 2016 semesters. Overall findings from this dissertation elucidated the presence of graphing competencies and difficulties in clinical and naturalistic settings in undergraduate biology students, graduate students, and professors, and informed the development and validation of three instructional and learning tools. These materials have the potential to resolve persistent difficulties with graphing and can be incorporated in teacher education and implemented in science classrooms at the undergraduate and K-12 levels.

CHAPTER 1. INTRODUCTION

1.1 Overview

There have been numerous calls at the national level to improve the undergraduate curriculum with respect to mathematics in STEM disciplines over the last 13 years. In 2003, the BIO2010 report recommended that undergraduate biology students use computer technology to collect and analyze data and to visualize data beyond simple bar and line graphs (NRC, 2003). In 2009, the Scientific Foundations for Future Physicians document listed "create and interpret appropriate graphical representations of data" and "make statistical inferences from data sets" as some of the expectations for students entering medical school (AAMC-HHMI Committee, 2009). Five years ago, in 2011, the Vision and Change in Undergraduate Biology Education document listed "ability to use quantitative reasoning" as one of the core competencies all undergraduate students should master, both in and outside the classroom (AAAS, 2011). These education policy documents over the past 13 years have emphasized the importance of engaging students with data and graphing as a skill and suggested instructors to teach this skill in their classrooms. In addition to these policy documents, Dirks and Knight (2016) compiled a list of essential biology concepts and competencies crucial for undergraduate students to master, with engaging in scientific inquiry and experimental design, and analyzing and evaluating data being 2 of the 6 essential competencies.

Graphs are a main component of the scientific language because they can be used to condense and summarize large datasets. The end result is a symbolic representation of experimental findings utilized by scientists for communication (Biechner, 1994; Tairab & Al-Naqbi, 2004; Wainer, 1992). The development of the skill to create appropriate and clear graphs is necessary for the scientifically literate individual (Padilla, McKenzie, Shaw, 1986; George & Bragg, 1996; AAMC-HHMI Committee, 2009; AAAS, 2011; Gormally et al., 2012). In order to facilitate this development and target instructional efforts, a full understanding of the skills and concepts relevant to graphing is required.

1.2 Research Aims of this Dissertation

The broad research aims of this dissertation are to: (1) understand student difficulties with graph choice and construction at the undergraduate level in both clinical and classroom settings and (2) develop and validate graphing materials to aid instructors and students when working with data, graph choice, and construction, with the goal of improving graph choice and construction.

1.3 Dissertation Chapters

This dissertation is divided into three parts: research and evaluation, curriculum development, and instruction (Redish, 2002). All studies in these parts were carried out under the constructivist learning paradigm, which states that individuals construct new ideas from their past knowledge and experiences (Duffy & Jonassen, 1992) The first framework under this paradigm is the expert-novice framework that explains how experts are able to perform tasks such as noticing meaningful patterns, how they embody a deep understanding in their field of study, how they integrate ideas, retrieve knowledge, approach new situations, and how experts disseminate their implicit knowledge (Chi, 2006; Brandsford, 2000; Postigo & Ponzo, 2004).

Embedded within the expert-novice framework are Meta-Representational Competence (MRC; diSessa & Sherin, 2000; diSessa, 2004) and the Cognitive Apprenticeship Model (CAM, Collins et al., 1991), which guided data analysis, material development, and recommendations to implement these materials in a classroom setting.

The research and evaluation chapters consist of chapters 2 and 3. Since we are operating under the constructivist paradigm, we utilized previous findings reported in the K-12 literature to understand the types of reasoning and potential difficulties experienced by students at this early age. Mevarech and Kramarsky (1997) and others have stressed the importance of understanding students' prior knowledge, since students construct their knowledge within their social and physical world, outside of the classroom, prior to formal instruction. Therefore, when students come into the classroom, they are bringing with them their prior set of beliefs and knowledge that may or may not agree with accepted meanings (Mevarech & Kramarsky, 1997). This gives us a starting point of what to expect and whether or not the same difficulties that are present in K-12 students persist in the undergraduate student population. For example, students are introduced to bar graphs and pictographs in the second grade (NGSS, 2012) in the context of mathematics. Throughout this dissertation, we refer to studies with graphing with elementary school children (Ainley, 1995; Aber-Bengtsson, 2006), middle school (Padilla, McKenzie, & Shaw, 1986; Wavering, 1989; Berg & Smith, 1994; Mevarech & Kramarsky, 1997; Kanari & Millar, 2004; Hattikudur et al., 2012), high school (Padilla, McKenzie, & Shaw, 1986; Wavering, 1989; Brasell & Rowe, 1993; Berg & Smith, 1994;

Ates & Stevens, 2003; Tairab & Al-Naqbi, 2004; Dori & Sasson, 2008), and undergraduate students (Picone et al., 2007; Bray-Speth et al., 2010; McFarland, 2010).

We also refer to the graphing literature with instructors (Roth & Bowen, 2001; Bowen & Roth, 2005; Roth, 2013), professionals (Rougier et al., 2014; Weissgerber et al., 2015), and medical doctors (Cooper, Schriger & Tashman, 2001; Schriger & Cooper, 2001; Schriger & Close, 2002; Schriger et al., 2006). Additionally we consult journal articles aimed to remediate graphing difficulties (Drummond & Tom, 2011; Duke et al., 2015; Saxon, 2015) and existing instructional books on graphing (Bertin, 1983; Tufte, 1983, Kosslyn, 1994; Few, 2004).

Our literature search revealed that students in grades 7-12 can successfully plot data points (Padilla, McKenzie, & Shaw, 1986), and determine X and Y coordinates of a data point (Padilla, McKenzie, & Shaw, 1986). These are not unexpected findings, since plotting data and determining coordinates are both skills emphasized starting in elementary school in both math and science courses (Padilla, McKenzie, & Shaw, 1986).

We also discovered that students have difficulties with the following: Graph choice (Leonard & Patterson, 2004; Tairab & Al Naqbi, 2010) labelling title and axes (Leonard & Patterson, 2004), understanding variables (Tairab & Al Naqbi, 2010), scaling axes (Padilla et al., 1986; Brasell & Rowe, 1993; Ainley, 1995), representing raw data accurately (Mevarech & Kramarsky, 1997; Bakker, 2004; Leonard & Patterson, 2004; Meletiou & Lee, 2010; Bray-Speth et al., 2010), understanding and plotting slope and y-intercept (Hattikudur et al., 2012) , performing simple calculations (Meletiou & Lee, 2010; Bray-Speth et al., 2010), misunderstanding best-fit line (Padilla, McKenzie, & Shaw, 1986; Brasell & Rowe, 1993), and problems with context (Hattikudur et al.,

2012) . Many of these identified difficulties highlight some mathematical concepts that are important for graph construction. However, constructing and interpreting graphs sits at the intersection of the mathematical, statistical, and discipline-specific (e.g. biology) concepts and skills with which students struggle.

The first part of the research and evaluation portion is chapter 2 of this dissertation, where we investigated the underlying step-wise reasoning that occurs when expert and novice biologists declare their graph choice and construct a graph to represent a biological dataset. This study is novel because of the dearth in the literature preceding this dissertation has not documented this phenomenon with expert and novice biologists. Findings from this study gave us a better understanding of what expert thinking looks like, how they perceive graphs and incorporate their mathematical knowledge into graphing. These data informed the development of graphing materials to help students with graph choice and construction. A semi-structured interview protocol was used to conduct think-aloud interviews with 5 professors, 8 graduate students (GS) and 15 undergraduate students, of whom 10 reported having no research experience (UGNR) and 5 reported having research experience (UGR). The participants used data from a biological scenario to construct their graph using a Livescribe pen, which captured the audio from the think-aloud and the pen strokes. Findings revealed that all professors planned and thought about data before graph construction. When reflecting on their graphs, professors and GS focused on the function of graph and experimental design while most UGR and UGNR relied on intuition and data provided in the task. Most UGR and UGNR meticulously plotted all data with scaled axes, while professors and some GS 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.

The second part of the research and evaluation portion is chapter 3 of this dissertation which reveals the reasoning implemented by undergraduate students enrolled in an upper-division physiology laboratory course when choosing and creating appropriate graphical representations of physiological data. Four times over the course of the spring 2013 and 2014 semesters, students (n=139) worked in small groups to design experiments relevant to that weeks' topic, collect data, and present findings in a short PowerPoint presentation. After the presentations, students were asked to individually reflect on graph choice, advantages, and disadvantages of their groups' graphical representation. At the end of the spring 2013 and 2014 semesters, student graphs were evaluated qualitatively based on four categories: graph mechanics, data form, graph choice, and aesthetics and visuo-spatial considerations, and the reflection responses were coded using thematic analysis (Boyatzis, 1998; Joffe & Yardley, 2004).

Findings reveal that the most common types of graphs constructed were: bar, scatter, or line graphs, and dot plots and box and whisker plots were the least common. Student reflections for graph choice fell into five main themes: *technology, graph interpretation, communication, reflection,* and *experimental concepts*. There was no explicit mention of the question or hypothesis in the students' reflections on graph choice. Students often correctly identified the advantages, associating variable type with graph type. However, when articulating the disadvantages, students often stated the advantages or self-critiqued their graphs. Findings from this study influenced the development of several graphing instructional materials to improve student reasoning with graph choice and construction.

Chapter 4 presents the curriculum component of this dissertation, where we share the development of three learning and instructional materials: Guide to data displays, step-by-step guide to data communication, and the graph rubric. Here, we highlight the purpose, development, validation, and usage of these materials, which increase students' knowledge with common graphs and provide a guiding framework for data presentation.

The last part of this dissertation is the instruction portion. In chapter 5, we explain how materials developed in chapter 4 were incorporated into an inquiry-based laboratory component of an upper division physiology course taken by biology majors at a large, Midwestern research-intensive university (n=123 students) during the Spring 2015 and 2016 semesters. In this class, students worked in small groups to design experiments, collect and analyze data, and present their findings to the class. Student learning of graphing was facilitated by applying the previously developed instructional resources (chapter 4) within the CAM components.

Student usage of the instructional materials were noted in the instructor's field notes, and the frequency of downloads of instructional materials was tracked from course management software throughout the semester. The effectiveness of the teaching intervention was evaluated by: pre/post survey on graph knowledge, the attributes and quality of student-generated graphs throughout the Spring 2015 and 2016 semesters, and comparison of students' graphs to those in the non-intervention semesters, Spring 2013 and 2014. Compared to the non-intervention semesters, the intervention semesters show that by the last lab, more student groups chose to construct either a box or dot plot and overall, the quality of graphs constructed was better.

There was no difference in graph choice in the pre and post surveys, as majority of the students chose line graph for their data display, however compared to the pre survey, more students in the post survey indicated that they would choose box and dot plots to display their data. This instructional approach, with its resources and practices, has provided further insights into difficulties that persist, which can be used to guide future instruction.

References

AAMC-HHMI Committee. (2009). Scientific foundations for future physicians. *Washington, DC: Association of American Medical Colleges*.

Dirks, C., & Knight, J. K. (2016). Measuring College Learning in Biology. *Improving Quality in American Higher Education: Learning Outcomes and Assessments for the 21st Century*, 225.

George, M. D., & Bragg, S. (1996). *Shaping the future: New expectations for undergraduate education in science, mathematics, engineering, and technology*. DIANE Publishing.

Åberg-Bengtsson, L., & Ottosson, T. (2006). What lies behind graphicacy? Relating students' results on a test of graphically represented quantitative information to formal academic achievement. *Journal of Research in Science Teaching*, *43*(1), 43-62.

Ainley, J. (1995). Re-viewing graphing: Traditional and intuitive approaches. For the Learning of Mathematics, 15(2), 10-16.

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

Ates, S., & Stevens, J. T. (2003). Teaching line graphs to tenth grade students having different cognitive developmental levels by using two different instructional modules. *Research in Science & Technological Education*, *21*(1), 55-66.

Bakker, A. (2004). Reasoning about shape as a pattern in variability. *Statistics Education Research Journal*, *3*(2), 64-83.

Beichner, R. J. (1994). Testing student interpretation of kinematics graphs. *American journal of Physics*, *62*(8), 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*(6), 527-554.

Bertin, J. (1983). Semiology of graphics: diagrams, networks, maps. Madison, WI: Univ Wisconsin Press.

Bowen, G. M., & Roth, W. M. (2005). Data and graph interpretation practices among preservice science teachers. *Journal of Research in Science Teaching*, *42*(10), 1063-1088.

Boyatzis, R. E. (1998). *Transforming qualitative information: Thematic analysis and code development*. Sage.

Brasell, H. M., & Rowe, M. B. (1993). Graphing skills among high school physics students. School Science and Mathematics, 93(2), 63-70.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How people learn: Brain, mind, experience, and school*. National Academy Press.

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*(3), 323-332.

Chi, M. T. (2006). Two approaches to the study of experts' characteristics. The Cambridge handbook of expertise and expert performance, 21-30.

Collins, A., Brown, J. S., & Holum, A. (1991). Cognitive apprenticeship: Making thinking visible. *American educator*, *15*(3), 6-11.

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*(3), 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*(1), 13-19.

diSessa, A.A., & Sherin, B. L. (2000). Meta-representation: An introduction. *The Journal of Mathematical Behavior*, *19*(4), 385-398.

diSessa, A. A. (2004). Metarepresentation: Native competence and targets for instruction. *Cognition and instruction*, *22*(3), 293-331.

Dori, Y. J., & Sasson, I. (2008). Chemical understanding and graphing skills in an honors case-based computerized chemistry laboratory environment: The value of bidirectional visual and textual representations. *Journal of Research in Science Teaching*, *45*(2), 219-250.

Drummond, G. B., & Tom, B. D. (2011). Presenting data: can you follow a recipe?. *Clinical and Experimental Pharmacology and Physiology*, *38*(12), 787-790

Duffy, T. M., & Jonassen, D. H. (1992). Constructivism: New implications for instructional technology. *Constructivism and the technology of instruction: A conversation*, 1-16.

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*(22), 3040-3059.

Few, S. (2004). *Show me the numbers: Designing tables and graphs to enlighten* (Vol. 1, No. 1). Oakland, CA: Analytics Press.

Gormally, C., Brickman, P., & Lutz, M. (2012). Developing a Test of Scientific Literacy Skills (TOSLS): measuring undergraduates’ evaluation of scientific information and arguments. *CBE-Life Sciences Education*, *11*(4), 364-377.

Hattikudur, S., Prather, R. W., Asquith, P., Alibali, M. W., Knuth, E. J., & Nathan, M. (2012). Constructing Graphical Representations: Middle Schoolers' Intuitions and Developing Knowledge About Slope and Y-intercept. *School science and mathematics*, *112*(4), 230-240.

Joffe, H., & Yardley, L. (2004). 4. Content and thematic analysis. *Research methods for clinical and health psychology. California: Sage*, 56-68.

Kanari, Z., & Millar, R. (2004). Reasoning from data: How students collect and interpret data in science investigations. *Journal of Research in Science Teaching*, *41*(7), 748-769.

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

Leonard, J. G., & Patterson, T. F. (2004). Simple computer graphing assignment becomes a lesson in critical thinking. NACTA journal, 48(2), 17-21.

McFarland, J. (2010). Teaching and Assessing Graphing Using Active Learning. *MathAMATYC Educator*, *1*(2), 32-39.

Meletiou-Mavrotheris, M., & Lee, C. (2010). Investigating college-level introductory statistics students' prior knowledge of graphing. *Canadian Journal of Science, Mathematics and Technology Education*, *10*(4), 339-355.

Mevarech, Z. R., & Kramarsky, B. (1997). From verbal descriptions to graphic representations: Stability and change in students' alternative conceptions. *Educational Studies in Mathematics*, *32*(3), 229-263.

National Research Council (US) Committee on Undergraduate Biology Education to Prepare Research Scientists for the 21st Century. (2003).*BIO2010: Transforming undergraduate education for future research biologists*. National Academies Press (US).

National Science Teachers Association. (2012). Next generation science standards. *Retrieved on*, *30*, 12.

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*(1), 20-26.

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.

Postigo, Y., & Pozo, J. I. (2004). On the road to graphicacy: The learning of graphical representation systems. *Educational Psychology*, *24*(5), 623-644.

Redish, E.F. (2002). Teaching with the Physics Suite. Accessed September 6th, 2016: http://www2.physics.umd.edu/~redish/Book/

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

Roth, W. M. (2013). Undoing decontextualization or how scientists come to understand their own data/graphs. *Science Education*, *97*(1), 80-112.

Rougier, N. P., Droettboom, M., & Bourne, P. E. (2014). Ten simple rules for better figures. *PLOS Comput Biol*, *10*(9), e1003833.

Saxon, E. (2015). Beyond bar charts. *BMC biology*, *13*(1), 60.

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*(1), 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*(6), 750-756.

Tairab, H. H., & Khalaf Al-Naqbi, A. K. (2004). How do secondary school science students interpret and construct scientific graphs?. *Journal of Biological Education*, *38*(3), 127-132.

Tufte, E. R., & Graves-Morris, P. R. (1983). *The visual display of quantitative information* (Vol. 2, No. 9). Cheshire, CT: Graphics press.

Wainer, H. (1992). Understanding graphs and tables. *Educational researcher*, *21*(1), 14-23.

Wavering, M. J. (1989). Logical reasoning necessary to make line graphs.*Journal of Research in Science Teaching*, *26*(5), 373-379.

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 Biol*, *13*(4), e1002128.

CHAPTER 2. EVALUATING META-REPRESENTATIONAL COMPETENCE WITH GRAPH CHOICE AND CONSTRUCTION ALONG THE NOVICE-EXPERT CONTINUUM IN BIOLOGY

2.1 Abstract

Undergraduate biology education reform aims to engage students in scientific practices such as experimentation and data analysis and communication. Graphs are ubiquitous in 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 meta-representational competence and expert-novice frameworks, we conducted think-aloud interviews to reveal differences in reasoning and graph quality between undergraduate biology students, with research experience (UGR), without research experience (UGNR), graduate students (GS), 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 GS focused on the function of graph and experimental design while most UGR and UGNR relied on intuition and data provided in the task. Most UGR and UGNR meticulously plotted all data with scaled axes, while professors and some GS 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.

2.2 Introduction

Graphs are the main components of the scientific language because they can be used to condense and summarize large datasets. The result is a symbolic representation that displays experimental findings utilized by scientists for communication (Biechner, 1994; Tairab & Al-Naqbi, 2004; Wainer, 2013). The development of the skill to create appropriate and clear graphs is necessary for the scientifically literate individual (Padilla, McKenzie, & Shaw, 1986). However, studies that document difficulties with graph construction date back 30 years. For example, researchers in the 1980's documented middle and high school student difficulties with line graphs: scaling axes, using a best-fit line, and assigning variables to axes (Padilla, McKenzie, & Shaw, 1986).

While there have been numerous suggestions to rectify difficulties with graph construction (e.g. sketch the graph before using software for construction (Patterson & Leonard, 2005), and incorporate more graphing opportunities into the classroom (Roth & McGinn,1997; Roth & Bowen, 2001; McFarland, 2010), data supporting the usefulness of these suggestions have not been documented in the literature. In addition, 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, graduate students, and active research scientists. Indeed, graphing difficulties exist and have been documented in individuals who possess advanced and/or terminal degrees, i.e. professors (Bowen & Roth, 2005), professionals (Rougier et al., 2014; Weissgerber et al., 2015), and medical doctors (Cooper et al., 2001; Schriger & Cooper, 2001; Cooper et al., 2002; Schriger et al., 2006). Although books (Bertin, 1983; Tufte, 1983; Kosslyn, 1994; Few, 2004) and short web-based instructional tools (Create A Graph Tutorial-NCES Kids' Zone, 2016; Interactive Statistics Map, 2016) exist to aid professionals (Rougier et al., 2014; Slutsky, 2014; Saxon, 2015; Weissgerber et al., 2015; Nuzzo, 2016) and students (Webber et al., 2014) as they create graphs, they provide superficial guidelines to the graph constructor on proper choice and construction of their graph. One contributing factor to this superficiality is that the reasoning that occurs during graphing has not been studied in depth. Questions remain, such as: What is the best way to sketch a graph before digitizing it using software? What should be considered when choosing a graph for data presentation? For instructors, what is the most effective way to teach students about proper graph choice and construction? What is useful feedback to provide students so they improve their graphing skills?

Answering these questions is the broader goal of our research, which is timely with the recent calls to reform the undergraduate curriculum with respect to incorporating aspects of data literacy into the STEM disciplines. 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. However, there have been increased efforts to formalize quality training for graduate students as instructors (Schussler et al., 2008; Reeves et al., 2016) and scholars (NSF Research Traineeship Program, 2016; NIH Institutional Training Grants, 2016). Within the discipline of biology, there is an emphasis to infuse quantitative reasoning into the classroom with an emphasis on creating and interpreting graphical representations (AAMC-HHMI

Committee, 2009; Brewer and Smith, 2011). The increasing implementation of Course-Based Undergraduate Research Experiences (CUREs) further signifies the importance of understanding how students grapple with data and data presentation to facilitate their mastery of this skill (See Figure 2.2.1 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, which may include data analysis and graphing (Bangera & Brownell, 2014; Brownell et al., 2015; Linn et. al, 2015). In order to answer national calls to infuse concepts relating to data and graphing into the undergraduate curriculum, we first need to understand areas of student competencies and difficulties with graphing, with a focus on graphing concepts and skills.

2.2.1 Areas of Student Difficulty with Graphing

Investigating teaching and learning of graphing in the context of undergraduate biology is rare (Bray-Speth et al., 2010; McFarland, 2010). As such, we review literature that precedes college and covers other science disciplines. These data are relevant because many students who enroll in an introductory biology course come from diverse backgrounds with a range of training and experience with data analysis and graphing.

Current trends in biology education place students in a position to engage in data analysis and graphing, however undergraduate students struggle with many fundamental concepts and skills relevant for graphing. 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 what the graph constructor intended. Extensive research has documented student difficulties with graph interpretation. Tairab and Al-Naqbi (2004) showed that students in tenth 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 & Janvier, 1992; Picone et al., 2007; Colon-Berlingeri & Borrowes, 2011).

While these studies focused on graph interpretation, the concepts and skills that they studied are an integral part of 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 & Smith, 1994; Friel & Bright, 1996; Grunwald & Hartman, 2010; Clase et al., 2010; Angra & Gardner, 2016). In order for a graph to be an effective communication piece for both the creator and the observer, there are four main components that should be considered: data form (Wild & Pfannkuch, 1999; Konold et al., 2015), graph choice (Cleveland, 1984; Schriger & Cooper, 2001; Metz, 2008; McFarland, 2010; Franzblau & Chung, 2012; Humphrey et al., 2014; Rougier et al., 2014), graph mechanics (Padilla, McKenzie, & Shaw, 1986; Kosslyn, 1994; Brasell & Rowe, 1993; Ainley, 2000; Leonard & Patterson, 2004; Bruno & Espinel, 2009; Bray- Speth et al., 2010; McFarland, 2010), and aesthetics and visuo-spatial considerations (Tufte, 1983; Kosslyn, 1994; Kostelnick, 1998; Kellman, 2000; Few, 2004). While there are four distinct components, they are all interrelated and influence the quality of the message communicated by the graph.

2.2.2 Data Form

The first, and perhaps most critical, step in graph construction begins with collecting, organizing, and understanding the data present. Wild and Pfannkuch (1999) coined the term transnumeration, which they define as “numeracy transformations made to facilitate understanding” (pg. 227). Translating a data table into a graph is one example of transnumeration. However, this phenomenon cannot exist alone. In order to produce a graph and to convey summaries of and trends in data, it is imperative for the graph maker to integrate their understanding of the raw data presented with their statistical and contextual knowledge actively and accurately (Wild and Pfannkuch, 1999). Konold et al., (2015) state that data can be viewed through multiple lenses: pointers, case values, classifiers, and aggregates. Each data lens can be used to organize data differently and to answer different questions (Konold et al., 2015). Although there is a hierarchy in these lenses (see Figure 2.3 in Konold et al., 2015) the aggregate data lens is the most complex and requires knowledge of statistics (descriptive or inferential) (Konold et al., 2015).

2.2.3 Graph Choice

With the advent of technology and the era of ‘big data’ which has necessitated the creation of new data visualizations using graphing software, the graph creator can now choose from a large array of graph types to represent data (Cleveland, 1984; Schriger & Cooper, 2001; Metz, 2008; McFarland, 2010; Franzblau & Chung, 2012; Humphrey et al., 2013; Rougier et al., 2014). Although the options for graph choice are expansive, one type of graphical representation is not necessarily better than another (Konold & Higgins, 2003). Each graph has a defined purpose, advantages, and disadvantages concerning the message that will be conveyed, the type of data displayed, and the conventions for graph mechanics (Bright & Friel, 1998; Schriger & Cooper, 2001; Konold & Higgins, 2003; Angra & Gardner, 2016). To the knowledgeable graph constructor, the plethora of graph choices and tools present options to optimize elements of the data display. However, the abundance of graphing options may cause confusion to the novice graph constructor (Grawemeyer & Cox, 2004), and incorrect graph choice can skew the take home message (Shah et al., 1999). Metz (2008) and Humphrey et al. (2014) have demonstrated student confusion between bar graphs and histograms. Some of this confusion can be attributed to the lack of consistency in the terminology between the two by textbook and web site authors, and another source of confusion stems from the lack of knowledge of variables (Leonard & Patterson, 2004; Bray- Speth et al., 2010; McFarland, 2010). Li and Shen (1992) report numerous examples of student graphs where inappropriate choices with graphs were made with respect to the variables students were given.

2.2.4 Graph Mechanics

All graphs constructed on the Cartesian coordinate system follow rule-bound conventions (Kostelnick, 1998). In this paper, we refer to these as graph mechanics (i.e. title, axes labels, units, scale, and key), which are elements that frame the data (Padilla, McKenzie, Shaw, 1986; Kosslyn, 1994; Bruno and Espinel, 2009). Literature reveals that students struggle with: formulating a descriptive title (McFarland, 2010), providing axes labels that are descriptive and match the variable type (Leonard and Patterson, 2004; Bray- Speth et al., 2010; McFarland, 2010), scaling axes (Padilla, McKenzie and Shaw, 1986; Brassell and Rowe, 1993; Ainley, 2000; McFarland, 2010), and constructing a key

(McFarland, 2010). McFarland (2010) illustrates difficulty with scaling axes by hand and using graphing software in undergraduate students. Li and Shen (1992) report similar difficulties with scale but using the computer software as the medium. They also report other deficiencies associated with scaling such as: omitting scales on one of the axes, omitting the zero value, failing to use axes breaks appropriately, and not providing sufficient divisions on the axes. Although having appropriate and detailed graph mechanics are required when communicating data to an audience, they are not always necessary when the graph constructor is trying to visualize data for their own understanding (Konold and Higgins, 2003). Therefore, knowing the purpose of the graph and who the graph reader will be prior to construction should determine the inclusion and quality of graph mechanics.

2.2.5 Aesthetics and Visual-spatial Considerations

A large portion of aesthetics is shaped by cultural knowledge and influences data design, which gives the graph its own visual voice (Kostelnick, 1998). Aesthetic approaches can reflect either contemporary taste or the designer’s intuition (Kostelnick, 1998). Many authors have advocated the modernist approach to graphing, adopting clean, minimalist designs that maximize data to ink ratio, while avoiding chart junk (Tufte, 1983; Kosslyn, 1994; Few, 2004). Graphs devoid of chart junk elements solely focus on the data that allows the audience to extract meaning from a graph (Tufte, 1983). By placing the focus on the data, two Gestalt laws of grouping, the laws of proximity and continuity, can be exploited to aid the viewer in extracting meaning from the graph (Kellman, 2000)

2.2.6 Suggestions to Remediate Graphing Difficulties

There have been numerous suggestions to remediate graphing difficulties. 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). Previous studies share sample datasets to encourage practice with graph creation (Patterson and Leonard, 2005; Tairab and Al-Naqbi, 2004; Bray-Speth et al., 2010).

In spite of the identification of the necessary components for graph construction and some research into difficulties with graphing and teaching interventions to remediate them, undergraduate biology students still struggle (Bray-Speth et al., 2010; McFarland, 2010; Hoffman et al. 2016). 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 numerous graphing choices that can be constructed, without having the graph constructor stop to evaluate and make important decisions like, “why am I constructing a graph?” “what is it that I'm trying to convey?” or “what variables will my graph show?”. In this study, we aim to uncover the reasoning that occurs during graph choice and construction and the attributes of the resulting graphs by utilizing the pen-and-paper mode of graph construction. Meta-representational competence and expert-novice frameworks as the lenses for this study guided data collection and analysis procedures.

2.3 Theoretical Frameworks Guiding Study Design and Analysis

The Meta-Representational Competence (MRC) framework outlines the knowledge and reflective reasoning practices that an individual competent in creating external representations, such as an expert scientist, would exhibit. As such, implicit in the MRC framework is expert-like knowledge and skill (diSessa, 2004) and using expert measures are helpful benchmarks when studying student MRC (diSessa, 2004; Bransford et al., 2000). Studying experts relative to novices illuminates differences between novices and experts and enables individuals who are less skilled at a task the opportunity to improve so that they reach expertise (Chi, 2006). Experts are assumed to be people who have acquired knowledge in a specific domain (Ericsson and Smith, 1991), constantly reflect on their thought process as they solve problems (Lyons, 2010), and organize their knowledge around big concepts and theories, which makes retrieving information easy and effortless (Bransford et al., 2000). Studying expertise in graphing is important because scientists have extensive experience with designing experiments, thinking about their data, producing and interpreting their own and others’ graphs, and are familiar with numerous types of graphs, and hence can shed light on what it takes to achieve competency with graphing (Roth and Bowen, 2003). Therefore, understanding experts is important because it provides insight into their adept knowledge and thought processes, which can inform classroom practices and inspire the movement of learning along the novice to expert continuum.

The components of the MRC framework can be leveraged to reveal a persons' areas of competence and difficulty with graph choice, construction, and critique (diSessa and Sherin, 2000). Specifically, these components are invention, critique, functioning, and learning or reflection (diSessa and Sherin, 2000). The first area, invention, reveals the underlying skills and abilities needed to conceive novel graphical representations from data (diSessa and Sherin, 2000). A common belief is that only expert scientists can create new representations (diSessa, 2004). However, students are equally capable of designing new representations and should be encouraged to engage with this as the first step to achieving competence and expertise (diSessa, 2004). The second area, critique, exposes the essential knowledge required to assess various types of graphs, their strengths and weaknesses (diSessa and Sherin, 2000). The third area, functioning, reveals the reasoning needed to understand the purpose of different types of graphs, with the usage being dependent on the type of data present (diSessa and Sherin, 2000). The final area, learning or reflection, reveals the awareness of a persons' own understanding of graphs (diSessa and Sherin, 2000).

The two frameworks described informed our study design, data collection, analysis, interpretation and conclusions. In addition, we use the MRC framework to define graph construction reasoning, one of the main targets of investigation in our study.

The MRC components of *invention* is assumed since everyone created a graph in our study. Therefore, we use the last three components from MRC to define graph construction reasoning as a persons' reflection on graph choice and construction by understanding the function of different types of graphs, being able to thoughtfully analyze a graph based on the type of data it is representing, variables, and the overall advantages and disadvantages represented by a 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 in order to gain automaticity and independence with graphing.

In this paper, we demonstrate, using semi-structured think-aloud interviews, the differences in graph construction reasoning between undergraduate students, graduate students, and expert professors when asked to use a simple data set to construct a graph. The interview was structured to capture the spontaneous thoughts of the participants as they underwent the task of graphing while also allowing the interviewer to probe for deeper knowledge and reasoning. Spontaneous reasoning appeared automatically during the graphing interview as participants worked on the graphing task. The guided reasoning occurred when the interviewer intervened and probed more deeply. Questions associated with the guided reasoning occurred after the participant finished their graph construction. See Appendix 2B for interview questions. Guided by the expert-novice framework, we divided the undergraduate students into two populations of students to explore differences that exist between students who have research experience and those who do not. Since students at the upper levels of their undergraduate education enroll in biology labs where collecting data and interpreting their findings in lab reports is mandated and have the most access to research apprenticeships in faculty laboratories, the criteria that we use to hierarchically differentiate between our novice students is based on research experience. Although research experiences are highly variable (Lopatto, 2007; Thiry et al., 2012), we find that by using this criteria, we can delineate the most novice students as those who do not have any research experience (UGNR), followed by the intermediate undergraduate students who have some research experience (UGR). In between the undergraduate students and professors, we have the graduate students who at the time of the study had at least completed their first year of graduate school. All graduate students had previous research experience from their undergraduate or graduate institutions. For complete demographic information, consult Appendix 2E. We include this population because it is a natural progression to professors and to the authors' knowledge, no literature regarding graph choice and construction exists on this population, making it a novel addition to our study. Lastly, we have professors, who are experts because they have more than 10 years of experiences constructing graphs (Chase and Simon, 1973; Ericsson, Krampe and Tesche-Romer, 1993; Ericsson, 2006).

2.4 Research Questions

The overarching research objective of this study is to elucidate the differences in graph construction reasoning that may exist between 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 without research experience (UGNR), with research experience (UGR), graduate students (GS), and professors reason with graph choice, data, and graph construction?
2. How do graph attributes differ between UGNR, UGR, GS, and professors?

2.5 Methodology

2.5.1 Development of the Graphing Scenario

The development of the scenario to be used in our study involved outside validation, literature review, and piloting. Knowing that at the time of the interview some of our participants would have had at most a partial semester of introductory biology, we consulted an award-winning high school teacher to get her opinion on scenarios that would be familiar to students who had ninth grade biology in high school. We decided to utilize two scenarios: bacteria growth or plant growth (Appendix 2A). Both bacteria and plant scenarios are isomorphic in that they both contained a dependent variable, independent variable, and two treatments with three replicates in each treatment. Simple numbers were used so that participants could easily manipulate the data, if they chose to do so (Konold et al., 2015). In four sentences, the scenario presented to the participant gave a brief background with a data table that organized all of the elements mentioned above. We chose to provide the participants with data organized 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 mentally the raw values (Wainer, 2013). While Garcia-Mila et al. (2014) saw differences in the types of graphs constructed by middle school students depending on whether the data were presented in a tabular format or list, this was not the aim of our work, so we presented our participants with data in a tabular format. The plant and bacteria scenarios were piloted with two undergraduate biology students and one professor to ensure readability and clarity. Pilot interviews were conducted in fall 2012 to solidify the interview protocol, prompts, and gauge the amount of time it took to construct a graph (Seidman, 2013).

2.5.2 Participant Recruitment

As part of a larger, multi-part 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). In order to obtain a heterogeneous and representative sample of the undergraduate student population, recruitment emails were sent to faculty teaching large biology courses. Personal emails were sent to graduate students and biology faculty requesting their participation in the study. All recruitment methods were approved by the Institutional Review Board (protocol no. 1210012775, see Appendix 2F). Recruitment criteria for undergraduate students were based on a) their status as or intention to be a biology major, and b) 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 emerged post-interview. Recruitment criteria for graduate students were based on a) their enrollment in the graduate program- all graduate students were pursuing a PhD degree; b) had successfully passed their qualifier examination, and c) had held a teaching assistantship or mentored undergraduate students. Criteria for professors were based on a) their credentials- all professors held a PhD. in a sub disciplinary field of biology; b) an active research laboratory with post-docs, graduate students, and/or undergraduate students; and c) having taught for at least one year.

2.5.3 Participants and Inclusion Criteria

Our initial pool of participants was 7 professors, 13 graduate students, and 39 undergraduate students. This pool was narrowed based on the following inclusion criteria. In order to minimize the threat to internal validity, we eliminated the 6 undergraduate and 1 graduate student interviews that were conducted early in the project with an interviewer who did not follow the semi-structured think-aloud protocol with high fidelity. Of the remaining 33 undergraduate student interviews that were conducted by the first author (AA), we further eliminated students who automatically constructed multiple graphs during the first prompt to construct a graph, as 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 the purpose of our study, we chose to exclude them to ensure uniformity of procedure and data. The same criteria were applied to graduate students and professors. Our final participant pool consisted of 5 professors, 8 graduate students and 15 undergraduate students. Of the 15 undergraduate students, 10 reported having no research experience and 5 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 lastly, the professors who each had more than ten years' experience conducting research and constructing graphs. Participants in our study represented many sub-disciplines in biology. Professors' specialties ranged from cellular neurobiology to behavioral ecology while the graduate students' research interests ranged from virology to avian behavior. Appendix 2E provides demographic information for our participants. Since 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.

2.5.4 Think-aloud Interviews for Graph Construction

For this study we used a pen-and-paper graphing task in the context of semi-structured, think-aloud interviews. The data reported here were collected from the first task in the multi-part interview and focused on understanding the reasoning behind graph choice and construction and the final graph artifacts. Although the complete think-aloud interview was long and ranged between 1-2 hours, the first part of the interview, which is the focus of this manuscript ranged only between 10-30 minutes. Prior to the think-aloud interview, participants were asked to complete an online survey through the Qualtrics software to gather their: demographic data (e.g. gender, race/ethnicity, course history/current enrollment, class standing, plans after graduation, student or professor status, and research experience or interests) and common ways they constructed graphs (i.e. computer, calculator, or by hand). To ensure quality motivation, participants were informed in the recruitment email and at the beginning of the interview that they will receive a $20 gift card for their time and efforts. Furthermore, we were transparent with our intentions that their responses will help us understand how professors and students create graphs with data, which will inform the development of instructional and learning tools to improve graphing. At that time, we asked our participants if they had any thoughts or questions about our research. We felt that establishing trust early and building rapport with each participant would allow us to collect quality data (Seidman, 2013).

All of the interviews were conducted between March 2013 and October 2014 in a soundproof interview room. The interview began with a few icebreaker questions and having the participant practice using the LiveScribe recording pen. Sample icebreaker questions are listed in Appendix 2B. The LiveScribe pen 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; Appendix 2A) predetermined before the interview. Each participant was asked to read scenario prompt aloud. They 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 they are performing a task. To account for this, the interviewer gently probed the participant 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 processes that occur in thought as well as 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 verses those who verbalized their thought (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 participant finished their graph construction, the interviewer intervened and asked the participant to reflect on their graph choice and data plotted. Interview questions can be found in Appendix 2B.

2.6 Data Analysis

2.6.1 Data Organization and Coding

In order to answer the first research question, think-aloud interviews were transcribed verbatim, and systematically organized and coded utilizing inductive analysis (Patton, 2005; Strauss and Corbin, 1998). The process of graph construction was divided into three phases of thought based on the problem-solving work of Polya (1945) and this was used to segment the transcript for coding. The three phases are planning, construction, and reflection. The planning phase is the duration of time that fell directly after the participant read the graphing task and before they drew their axes for the graph. The construction phase began with the drawing of the axes and ended when the participant signaled that they finished constructing their graph. Lastly, the reflection phase is when the interviewer intervened and probed the participant to elaborate on their graph and data choices. This initial step of transcript segmentation began the process of open coding within each phase of thought. 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. Two researchers independently coded at each step of the coding process, met regularly to compare and discuss the coding, until a consensus was reached on the final themes.

Due to the small sample size in each participant group, statistics for themes are not reported, but the absence or presence of themes along with the occurrence of the MRC categories between the three participant groups are summarized in Figure 2.5. Figures 2.2-2.4 illustrate themes present during the interview for each phase in the form of a gray-scale heat map in order to view patterns quickly in the data from multiple perspectives. Each participant is represented as a row while a column represents each theme. The prevalence of codes under each theme is conveyed by the shade of gray (light gray (least prevalent) to dark gray (most prevalent)). The heat map allows us to make quick data comparisons within and across themes and participant groups.

Transcript length showed that professors usually talked longer than the undergraduate students did in the planning and reflection phases (Figure 2.2.1). To see if there was a difference amongst the participant groups in terms of the time it took to plan, construct, and reflect on the graph, an independent samples t-test was conducted using Statistical Package for the Social Sciences (SPSS V.22, 2013). Levene’s Test for the equality of variance was conducted and when reporting the p value ($\alpha<.05$), equal variances were not assumed (SPSS V.22, 2013). Since we are interested in differences across participant groups, we did not perform inferential statistics across phases of the graph interview. Professors also used more words than the 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. In order to standardize time spent talking by each participant, word analysis was performed. Words mentioned multiple times within in a given phase were counted and coded once. The number of words for each code was counted and divided by the total 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 Figures 2.2-2.4.

To address the second research question, graphs constructed by professors, graduate students, and the two undergraduate population groups were described qualitatively based on four broad categories: graph mechanics, data form, graph choice, and aesthetics. The detailed list of evaluation categories are listed in Appendix 2C.

2.7 Results

We employed qualitative methods with pen-and-paper graphing think-aloud interviews to gather and analyze data to answer our two research questions:

1. How do undergraduate students without research experience (UGNR), with research experience (UGR), graduate students (GS), and professors reason with graph choice, data, and graph construction?
2. How do graph attributes differ between UGNR, UGR, GS, and professors?

Themes that emerged from the transcripts from our think-aloud graph construction interview are presented for each phase of the graph construction process (planning, construction, and reflection). In addition, we present a qualitative analysis of the graphs constructed by our participant groups.

2.7.1 Planning Phase

The planning phase occurred after the participant was presented with the task and before they began graph construction, as indicated by the drawing of the axes. There are many decisions that must be made prior to graph construction to ensure graph quality and clarity of the message being communicated. Within the planning phase, we expected participants to make sense of the data table and scenario, reflect on the purpose of the graph, think about ways to organize the data on the graph, manipulate the data, categorize the data and choose a graph type (Friel and Bright, 1996; Ainley, Nardi and Pratt, 2000; Patterson and Leonard, 2005). Figure 2.2.1 displays the amount of time the participants spent talking in each of the interview 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 least amount of time planning compared to the other phases.

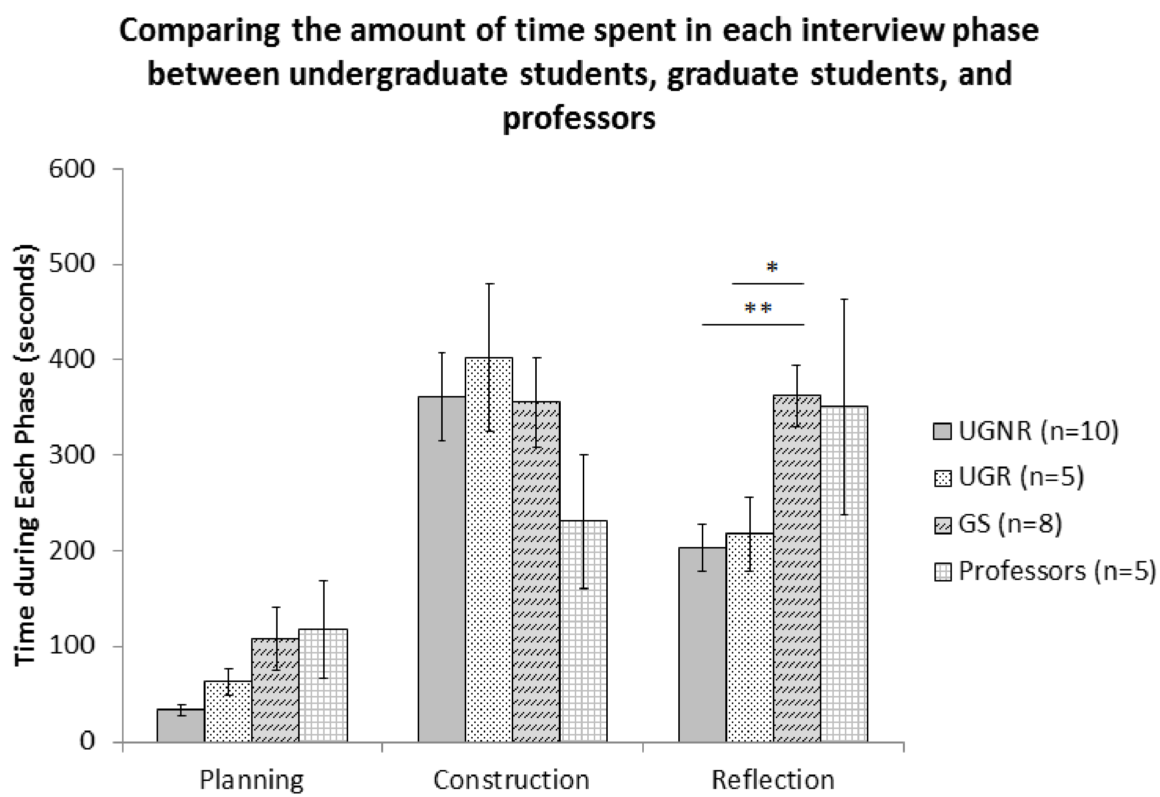


Figure 2.1: Comparing the amount of time spent in each interview phase between undergraduate students, graduate students, and professors: Summary of the time spent during the planning, construction, and reflection phases for professors (P), graduate students (GS), undergraduates with research experience (UGR), and undergraduates without research experience (UGNR). An independent samples t-test shows there was a significant difference in the amount of time spent reflecting between GS and UGR (*$p<.05$) and GS and UGNR (**$p<.01$;)

In order to characterize the reasoning used in the planning phase we mapped themes that emerged from our coding to the categories of the MRC framework. There are three categories from the MRC framework that apply to the planning phase: function, reflection, and invention (Table 2.1). The five themes that emerged from the participants during the planning phase were: *purpose, graph choice, data type, data table,* and *graph construction*. The definitions of the themes, example quotes, and the alignment of the themes to the MRC categories can be found in Table 2.1. Figure 2.2 displays the themes that emerged for each individual subject in the planning phase. These data are visualized in the form of a gray-scale heat map in order to view the patterns in the data from multiple perspectives. First, looking within a participant group but across themes, we see that professors and GS were more varied in their codes. This is indicated by the presence of the various shades of gray across multiple themes. Second, looking across participants but within a theme, we see that regardless of participant group, *data table* was the most prevalent theme and *purpose* was the least prevalent theme.

Table 2.1: Planning Phase: Summary of the themes, definitions, and participant examples

Categories in MRC	Themes	Participant Examples
Function- Providing reasoning for understanding the purpose of different representations, their usage, and limitations (diSessa & Sherin, 2000)	**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 do the 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 -explicitly stating graph choice (ie. 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* *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 [...] yeah 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 here like showing a time course of a single plant.* *UGNR4: I'm going to think that I'm going to make a line graph to compare two different types of data in the same graph [...] I think it's gonna show best the patterns of each.*
Invention- The underlying skills and abilities that allow students to conceive novel representations (diSessa & Sherin, 2000)	**Graph Construction** -when the participant either verbalizes variables in the table to the axes on the graph or the data or how 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: 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: when you have time you want to put that on the x axis*
	Data Type- explicitly making decisions about whether or not to plot raw data or plot manipulated data (ie. average) or the number of graphs to use to properly convey the data.	*P5: So let's do first on the 15 mL and what I would do is, since we have 3 times per time point I will try to get the average of the three.* *GS1: I would probably pool the replicates although the math for this would be pretty bad in my head and I would have to draw fake error bars.* *UGNR1: Okay um well I might be able to make this into two graphs just because it'll be easier to see maybe? Or we could do the average of the three tubes.*
Learning or Reflection- Strategies for fostering understanding of representations (diSessa & Sherin, 2000)	**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.* *GS6: I'm looking at the time, and to the number of cells, three test tubes.* *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.*

All five professors planned before constructing their graph (Figure 2.2). Four out of the five professors focused on understanding the data table and making decisions on the type of data and the number of graphs to use to represent the data best according to the scenario and purpose of the task. Professors were unique in that they were the only group who did not explicitly state the graph choice in the planning phase. Professor 4 did mention that they could visualize the data in *"two different colored lines"* but did not verbally finalize their decision in the planning phase (Refer to Table 2.1 for examples of direct quotes).

Seven of the eight graduate students planned before constructing their graph (Figure 2.2). Five of the eight graduate students spent their planning phase talking about the *data table*. Similar to the professors, graduate students verbally articulated how the variables in the data table would translate to the final graph form. However, unlike the professors, the graduate students did not state the *purpose* of the graph while they were planning. Two graduate students did articulate their *graph choice*, stating they could either make a line graph (GS6) or a scatter plot (GS4). Two graduate students reflected on the data, specifically thinking about calculating averages (GS1 and 6). Five graduate students mentally started to plot data in a graph. (Refer to Table 2.1 for examples of direct quotes.)

Of the five UGR, four individuals planned before proceeding to the construction phase (Figure 2.2). Participant UGR4 jumped straight into graph construction after reading the scenario, uttering *"Okay, so well always start with the axes"*. Of the four participants who did plan, three of them verbalized and reflected on the data provided in the table. Similar to the professors, UGR2 and UGR5 articulated the *purpose* of the graph and UGR1 and UGR 5 articulated how they visualize the data on the graph. Unlike the professors but similar to some of the graduate students, both UGR1 and UGR3 verbally articulated their *graph choice*. After initially deciding on a bar graph, UGR1 changed to a line graph because they wanted to represent the variable time (Refer to Table 2.1 for examples of direct quotes.)

Of the ten UGNR, seven individuals planned before proceeding to the construction phase (Figure 2.2). Participants UGNR2, UGNR3, and UGNR10 proceeded directly to graph construction. Similar to the professors and UGR, of the seven participants who planned, five verbally described the *data table*. This is the only theme that was present for UGNR5, 6 and 8. UGNR1 was the only participant to articulate the number of graphs they felt were needed to convey the data properly. UGNR4 was the only participant in this group to state their *graph choice* explicitly. (Refer to Table 2.1 for examples of direct quotes).

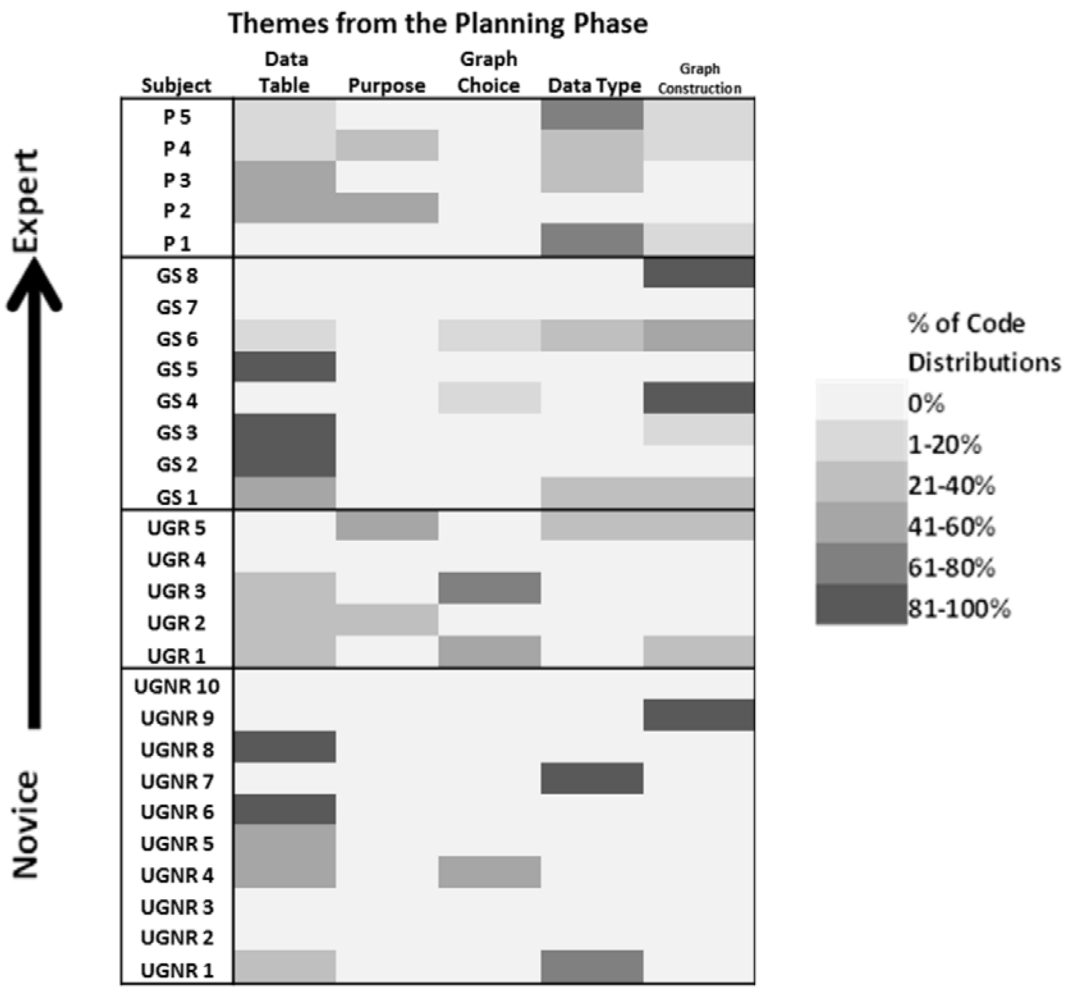


Figure 2.2: Planning Phase Themes: Heat map summarizing the percentage of code distributions present in each of the five themes present during the automatic planning phase for professors (P), graduate students (GS), undergraduates with research experience (UGR), and undergraduates without research experience (UGNR).

2.7.2 Construction Phase

The construction phase followed the planning phase and began with the drawing of the axes and ended when the participant signaled that they finished their graph construction. In order to properly construct and communicate a clear take-home message with a graph, the graph constructor requires basic knowledge of the different types of graphs, their usage with appropriate variables, advantages, disadvantages, graph mechanics (e.g. proper axis labeling) and aesthetics (Bertin, 1983; Kosslyn 1994; Few 2004). The graph constructor also needs to decide the purpose of the graph and how to display the data in the graph (Ainley, Nardi and Pratt, 2000; Patterson and Leonard, 2005). These decisions are not always practiced when graphing software are used (Ainley, Nardi and Pratt, 2000; Patterson and Leonard, 2005). Participants in our study created graphs using the LiveScribe pen-and-paper method, affording them the opportunity to not simply click a button to create their graphs, but to create their graphs deliberately. Relative to the planning phase, most participants spent more time constructing their graph (Figure 2.1). However, professors spent less time than the other three participant groups. This is consistent with the graphs they created (see Graph Attributes).

All four MRC categories were present in the construction phase, with a targeted focus on invention. Ideally, as participants are constructing their graph, they should also reflect on their graph choice, critique the data provided, 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 2.2. Figure 2.3 displays the seven themes that emerged from this phase across the participants: *data type, graph choice, aesthetics,* *technology, evaluation, sample size*, and *statistics*. First, looking within a participant group but across themes, we see that professors were the least varied in their codes. This is indicated by the lack of various shades of gray across multiple themes. Second, looking across participants but within a theme, we see that regardless of participant group, *data type* and *evaluation* were the most prevalent themes and aesthetics and technology were the least prevalent themes.

All five professors verbally articulated their thoughts aloud while constructing their graph. Professor 1 manipulated data and plotted the summation of bacteria cells at each temperature and thus their codes were focused within the theme of *data table*. Since Professor 3 evaluated and reflected on how to display data as they constructed their graph, their codes fell into the theme of *evaluation*. Professor 3 was the only participant who relied on their microbiology knowledge and mentioned that the data could be plotted on a logarithmic scale. Professors 3, 4, and 5 mentioned descriptive *statistics*, particularly adding error bars around each averaged data point to indicate the variation. Professors 2 and 4 critiqued the data presented and indicated that a bigger *sample size* would be preferable in order to run inferential statistics. However, professor 4 connected the small *sample size* to a possible real-life situation that 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*." Examples of direct quotations are presented in Table 2.2.

Table 2.2: Construction Phase: Summary of the themes, definitions, and participant examples

Categories in MRC	Themes	Participant Examples
Function- Providing reasoning for understanding the purpose of different representations, their usage, and limitations (DiSessa & Sherin 2000)	**Graph Choice** - explicitly stating graph choice (ie. 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. Cause the time line can be discrete. I'm not sure. I think since its cell growth over time that should be fine so.* *UGR2: And I'm just using the line graph because it kind of shows the trend the easiest because it goes straight and it goes up a little* *UGR3: is this line did I say line or bar? I'm doing lines. I'm doing a line chart now I changed my mind.*
Invention- The underlying skills and abilities that allow students to conceive novel representations (DiSessa & Sherin 2000)	**Statistics** - Talking about either descriptive or inferential statistics	*P5: and exact cause there is some error which I didn't calculate (draws error bars on each data point at 7:37 min) [...] and there will be some error* *P2: but the error for each one and then so that would be for the 10 (labels line as 10 C). And then I will use something different for the 22 (draws the line; labels as 22C). And also the error bars* *GS7: show the standard deviation from the average value* *UGR4: you can create a trendline for each dataset so basically out of 15ml and 5ml, you can do the line of best fit, where is you just try and really roughly like 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 just so you can tell where it's going.*
	Data Type- explicitly making decisions about whether or not to plot raw data or plot manipulated data (ie. average), and the number of graphs to use to properly convey the data	*P1: I'm collapsing across tubes so I'm giving total or I could do means* *GS5: There are three tubes so I think I need to within each temperature group I will do the average- calculate the mean of the number of cells for the same time point for all three tubes.* *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: Okay so, because we have three plants, which is like three trials for each, I'm gonna 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- Strategies for fostering understanding of representations (DiSessa & Sherin 2000)	**Evaluation**- Talking either about the general graphing habits, future directions, or take home message	*P2: but the error for each one and then so that would be for the 10 (labels line as 10 C). and then I will use something different for the 22 (draws the line; labels as 22C). and also the error bars* *GS8: This is the most horrible graph ever. Because it's not even clear what the data mean.* *UGR4: So you can see really clearly that they are increasing at the same rate but throughout the entire experiment, the 5ml produces less leaves* *UGNR3: I did this wrong [...] I should have put ml on the y axis [...] Um (5 second pause) I'll just keep going with this. I might be I might be okay. two (5 second pause) okay yeah I need to plot this with number of leaves instead of ml. (scratches the x axis label and re names is number of leaves) so the number of leaves will be on the x axis*

Table 2.2 continues

Categories in MRC	Themes	Participant Examples
Learning/Reflection- Strategies for fostering understanding of representations (DiSessa & Sherin 2000)	**Technology**- Mentioning the habitual graph making software to reflect on elements of their graph construction.	*GS3: So if I read the problem and use excel, it can just put linear regression in each case And I will put the r2 values both are greater than 0.8 or something* *UGR1: So I feel like in excel if I was doing this, I would make each plant like its own representation symbol or its own color to better represent that. Have like a uniform structure to this but a different* *UGR4: if you are in excel, you can say give me the equation for the trend line and it will tell you that y equals some function of x, so from that you can see the mathematical relationship behind the number of leaves that you have*
Critique- Critical knowledge that is essential for assessing the quality of representations (DiSessa & Sherin 2000)	**Aesthetics**- Using elements of graph design (i.e. gestalt principles and color) to critique the constructed graph.	*UGR2: I guess I will graph the other ones too and we can just imagine that they are different colors* *UGNR1: I'd use different colors like different colors for the ones at 22 and the ones at 10 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- Critiquing the small sample size presented in the data table.	*P4: With 3 plants in each, I guess you could put at standard error on that, n=3 is pretty small but sometimes in biology, you are stuck with pretty small. And I can't do that in my head but what I would probably do is put each of these, each point, plus or minus the standard error* *P2: You do need a bigger sample size* *UGNR6: if there was more data it could possibly be curving off to give a constant average at least if you want any of those*

Seven of the eight graduate students articulated their thoughts aloud while constructing the graph (Figure 2.3). Six graduate students reflected on *data type*. A new theme that emerged in the graduate student group was *graph choice*, seen with GS 1, 4, 7, and 8. GS1 talked about whether or not to present the data as a scatter plot or line graph from the following quote: "*Oh that's a good point, whether or not I can connect them. Cause the time line can be discrete. I'm not sure. I think since its cell growth over time that should be fine so.*" Another theme that emerged with the graduate student group was *technology*. GS3 mentions using excel to display linear regression and corresponding R^2 values: "*So if I read the problem and use excel, it can just put linear regression in each case and I will put the R^2 values both are greater than 0.8 or something.* Similar to the expert professor group, the theme *statistics* emerged four times with GS 1, 3, 5, and 7, where they mentioned adding error bars around the average data point. The theme *evaluation* emerged four times, with GS8 reflecting on their general graphing habits, GS3 reflecting on the data, and GS 1, 3, and 5 formulating the take-home message. Examples of direct quotations are presented in Table 2.2.

All five UGR articulated their thoughts aloud while constructing their graph (Figure 2.3). UGR1, 2, 3, and 5 had two themes each, whereas UGR4 had 4 themes and was the most similar to GS3. UGR1 spent the majority of the construction phase referring to the Microsoft Excel software package for graph construction and referring to *data type*. The codes from UGR2 were distributed equally between two themes: *graph choice* and *aesthetics*. For graph choice, UGR2 stated the reasoning for choosing a line graph was to show the trend of increasing growth. For the aesthetics, the reasoning was to imagine the use of color so that the separate data lines would be easily differentiated from each other.

For UGR3, the themes of *data type* and *graph choice* emerged. UGNR3 plotted some raw data and although the reasoning was not explicitly articulated, UGR3 changed their mind for the graph from bar to line; they insisted that a line graph is better. This is explored in the next phase of graph reflection. Lastly, UGR5 spoke about the type of data they could plot, ultimately deciding to calculate the averages. UGR5 also concluded their graph construction by formulating a take-home message. Examples of direct quotations are presented in Table 2.2.

Of the ten UGNR, seven participants articulated their thinking and decision-making during graph construction. This stood out from the other three participants in this group whose verbal narration 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. Of the seven participants, only four participants who did articulate their reasoning, had only one theme emerge, each. For UGNR1, the theme of *aesthetics* emerged, where they articulated using different colors to differentiate between each treatment group. We also saw this theme emerge for UGNR7, who talk about the inability to discern certain data points from others. For UGNR3, the theme that was present was evaluation and although this participant was talking aloud about the mechanics that encompass graph construction. They also spent the entirety of the construction phase confused because they switched the axes. An excerpt from UGNR3:

"I did this wrong. […] I should have put ml on the y-axis. […] Um (5 second pause) I'll just keep going with this. I might be I might be okay. two (5 second pause) okay yeah I need to plot this with number of leaves instead of ml. (scratches the x-axis label and re names is number of leaves) so the number of leaves will be on the x-axis. […] Because we

are plotting leaves vs. time which I didn't realize at first and I need to switch the key to plants 1, 2, and 3 (makes new key)"

One possible explanation for the confusion of variables and plotting data can be attributed to the fact that this participant did not have a planning phase and jumped straight into graph construction (See Figure 2.2 and Discussion). For UGNR 8, the theme *data type* emerged, where they reflected on the number of graphs to use to display all of the different trials adequately. For UGNR9, the theme *evaluation* emerged, where they spent time articulating the take-home message. It is interesting to note that UGNR 6, 7, and 10's reasoning during graph construction was more advanced than the other students in this category. UGNR6 was the only student who mentioned *sample size,* a theme that did not emerge in either the GS or UGR groups. UGNR 7 and 10 reflected on their *graph choice*, deciding on a line graph. UGNR 6 and 7 were also the only participants who spoke about plotting a line of best fit to see if the trend was linear.

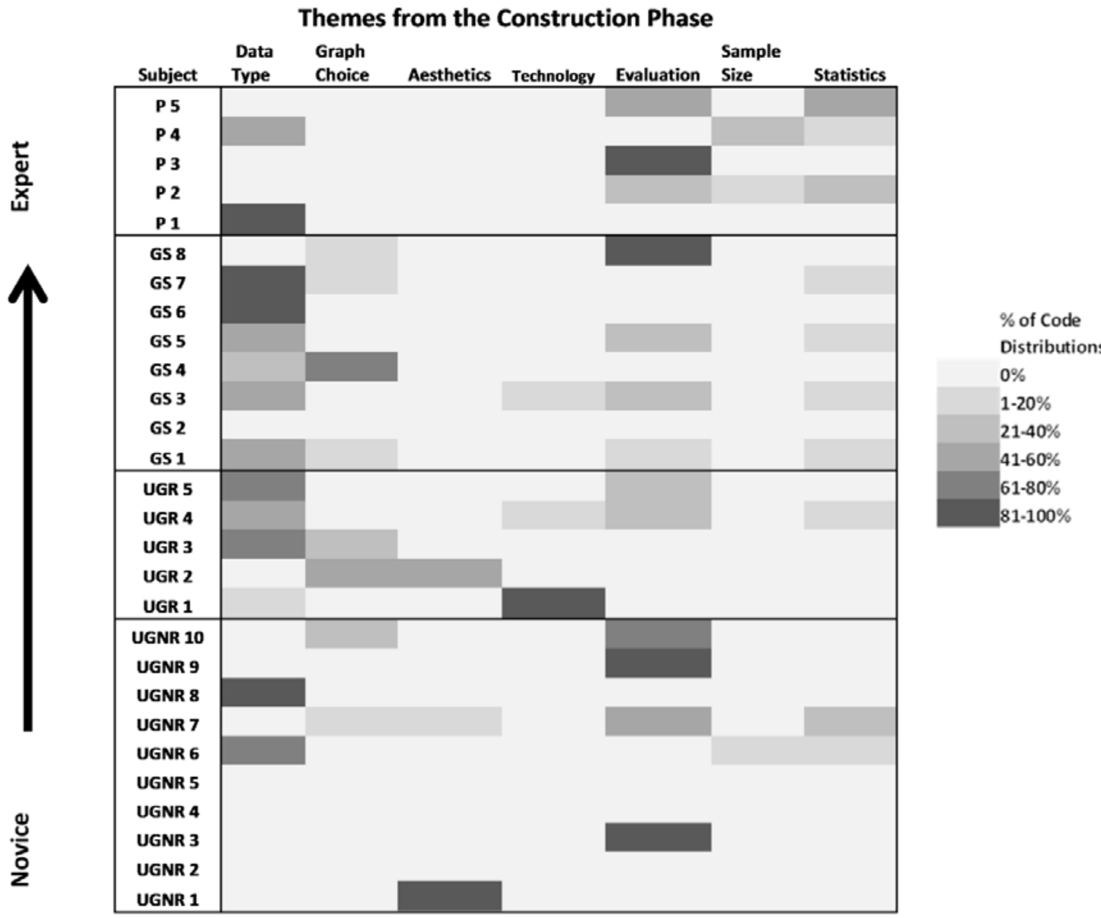


Figure 2.3: Construction Phase Themes: Heat map summarizing the percentage of code distributions present in each of the seven themes present during the automatic graph construction phase for professors (P), graduate students (GS), undergraduates with research experience (UGR), and undergraduates without research experience (UGNR).

2.7.3 Reflection Phase, Graph Choice

The reflection phase followed the construction phase and began when the interviewer intervened and probed the participant to elaborate on their graph choice (see Appendix 2B for interview questions). All four MRC categories were present in the reflection phase, targeting the learning and reflection category. We expected participants to elaborate on graph choice, using their graph created in the construction phase (invention), provide a reflection and critique. Figure 2.1 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 GS and UGR ($p<.05$; independent samples t-test, SPSS, V.22) and GS and UGNR ($p<.01$; independent samples t-test, SPSS, V.22)

A summary of the MRC categories, themes, and examples from transcripts are displayed in Table 2.3. All participants provided an answer for this phase. Figure 2.4 summarizes the seven themes that emerged from this phase: *statistics, evaluation, data table, variables, aesthetics, purpose,* and *time*. First, looking within a participant group but across themes, we see that UGNR were the least varied in their codes, indicated by the lack of various shades of gray across multiple themes. It is interesting to note that UGNR only talked about *statistics, evaluation,* and *data table* during the reflection phase. Second, looking across participants but within a theme, we see that regardless of participant group, *evaluation* was the most prevalent theme, which is not surprising since the participants were probed to reflect on their graph choice. The themes *data table, variables,* and *aesthetics* were the least common across the participant groups.

Table 2.3: Reflection Phase: Summary of the themes, definitions, and participant examples

Categories in MRC	Themes	Participant Examples
Function- Providing reasoning for understanding the purpose of different representations, their usage, and limitations (diSessa & Sherin 2000)	**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 wanna see the effect of temperature on growth, here you can easily see the two treatments, two levels of temperatures that were used while they changed over time* *GS4: because my question was how temperature affects the growth of bacteria.*
	Time- Using words like, change over time, or flow over time to justify choosing a line graph	*P5: I would say that most of the time for time it's the line graph* *GS1: I would be able to show how the cell number would change over time* *UGR1: hang on uh I mean things uh things that are measuring like changes over time I think lines show trends there better than my initial thought of a bar graph*
	Variables- 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: Because I was trying to decide whether or not time was going to be a continuous variable*
Invention- The underlying skills and abilities that allow students to conceive novel representations (diSessa & Sherin 2000)	**Statistics**- Talking about either descriptive or inferential statistics	*P2:but there could be an interaction between time and temperature and here you can see there doesn't appear to be one* *GS3: in the beginning I was thinking putting the standard deviation but I realized that I mean after I mean I decided to put that data first so kind of a line in here I think that putting a linear regression that is very easy to use and read.* *UGR4: Cuz you can't like compare the number of leaves for 15 mls at like after 120 hours with like the 5ml at 30 hours because that's just not a fair comparison, so you have to be able 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 not quite they're a little bit off which could be due to like 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- Strategies for fostering understanding of representations (diSessa & Sherin 2000)	**Evaluation**- Talking either about the general graphing habits, future directions, or take home message	*P5: if this would be 4 different plants instead of time points then I probably would have used a bar graph, if there would be more categories* *GS8: I were to do any other type of bar graph or something, I'm, 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: because one of the scales in the experiment was the passing of time, that's like, you can't use like a bar graph or pie chart to show that, to show the passing of time, cuz you're gonna want to show it like linearly along some kind of axis, so that means either like um you're gonna have to find some way to put the data points sequentially according to the time it happened, in order to be able to compare them accurately* *UGNR1: Um probably because this is the most common type of graph that I make so I thought of this kind first*

Table 2.3 continues

Categories in MRC	Themes	Participant Examples
Learning/Reflection- Strategies for fostering understanding of representations (diSessa & Sherin 2000)	**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 kind of* *UGR1: Cause the way this chart is presented, at first I kinda 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 um time vs. number of leaves, you'd have to do a scatter plot of sorts and in retrospect I should have made two graph and separated them out into 5 and 15ml* *UGNR1: Because that's what I thought about when I first looked at this chart and (3 second pause) it does show that the number of cells*
Critique- Critical knowledge that is essential for assessing the quality of representations (diSessa & Sherin 2000)	**Aesthetics**- 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: um yeah like I said ideally, this would be a little bit more visually appealing with different colors and you know and evenly spaced dots and lines*

During the reflection phase, five themes emerged from the expert professor population: *purpose*, *evaluation*, *variables*, *time*, and *statistics* (Figure 2.4). Professors 1 and 3 *evaluated* their graph and offered suggestions for other graph types that could have been used, or things they would do differently in the future. Professors 2 and 4 justified their graph choice by stating the purpose of the experiment and aligning it with the message portrayed by their graph (Figure 2.4). Professors 4 and 5, used time as one of the themes to justify their choice of a line graph. Direct quotations are displayed in Table 2.3.

For the reflection phase for graduate students, all seven themes were present. Like the expert professor group, GS 4 and 7 reflected on the *purpose* of the graph. GS 1 and 4 talked about *variables*. *Evaluation* was a popular theme amongst the graduate students and responses ranged from talking about personal experiences (GS 2, 3, 6, and 8), other graph types (GS7 and 8), and formulating the take home message (GS 4). GS1 was the only participant in the group to refer to the *data table* when answering the question and also the only one to link *time* to their line graph. GS3 was also the only participant to talk about *statistics* in their reflection. GS 8 was the only participant to talk about the *aesthetics* as they reflected on their graph choice.

From the UGR population, five themes emerged: *statistics, evaluation, data table, aesthetics*, and *time*. As with the expert professors, UGR 3 and UGR1 also used *time* as one of the themes to justify their choice of a line graph. UGR1 was the only participant to mention *aesthetics* stating that their graph would be more visually appealing if different colors were present and if the data points were evenly spaced apart. UGR1 was the only participant to use the *data table* to justify their graph choice. Three participants, UGR1, 2, and 4 used either inferential or descriptive *statistics* to explain their graph choice. All participants provided some sort of *evaluation* for their graph. UGR4 provided the most elaborate *evaluation* that resembled that of the expert professors, reflecting on other graph types, the take home message, and personal experiences. UGR5 also *evaluated* on the graph, but used their personal experiences and intuition to explain their graph choice. Direct quotations are displayed in Table 2.3.

Only three themes emerged from the UGNR population during the reflection phase: *evaluation*, *data table,* and *statistics*. While *evaluation* was a common category for 9 of the 10 participants, it consisted of various ideas.   For instance, UGNR1, 2, 8, and 9 relied on their personal experiences and intuitions when reflecting on their graph choice. UGNR4 and UGNR 10 on the other hand, spent their time formulating the take home message for the graph. UGNR5 and UGNR 7 was also different in that they justified their graph choice by explaining why bar, pie, and scatter plots would not accurately display the data. UGNR1 and 3 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 UGR1 and GS1. Please refer to Table 2.3 for examples of direct quotes.

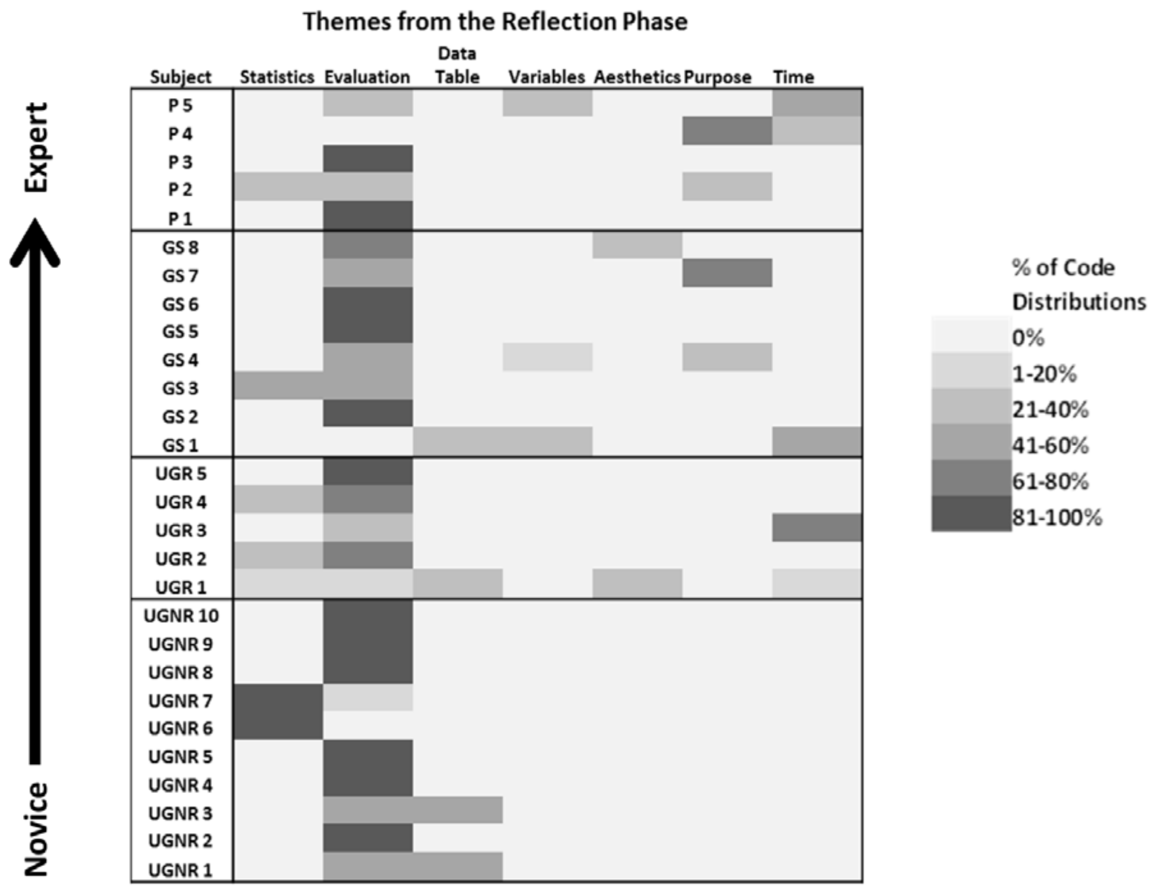


Figure 2.4: Reflection Phase Themes: Heat map summarizing the percentage of code distributions present in each of the seven themes present during the reflection phase on graph choice for professors (P), graduate students (GS), undergraduates with research experience (UGR), and undergraduates without research experience (UGNR).

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)	**X**		**X**			**X**					**X**	
	GS (n=7)	**X**		**X**						**X**		**X**	
	UGR (n=4)	X		**X**			**X**			**X**		**X**	
	UGNR (n=7)	**X**		X						X		**X**	
Construction Phase	P (n=4)	**X**	**X**		**X**								**X**
	GS (n=7)	**X**	**X**							**X**	**X**		**X**
	UGR (n=5)	**X**	X			X				**X**	**X**		**X**
	UGNR (n=7)	**X**	**X**		X	**X**				**X**			**X**
Reflection Phase-Graph Choice	P (n=5)		X				**X**	X	**X**				**X**
	GS (n=8)		X			X	**X**	**X**				X	**X**
	UGR (n=5)		**X**			X			**X**			X	**X**
	UGNR (n=10)		**X**									**X**	**X**

Figure 2.5: Summary of graph construction reasoning findings -Summary of the presence of themes in each of the three interview phases by professors (P, n=5), graduate students (n=8), undergraduates with research experience (UGR, n=5), and undergraduates without research experience (UGNR, n=10). An ‘X’ denotes the presence of a theme by one participant; An ‘X’ indicates the presence of a theme by multiple participants. Since invention involves graph construction and participants were explicitly asked to reflect on graph choice, these themes are blacked out. Refer to figures 2, 3, 4 for themes that appeared for each participant.

2.7.4 Graph Attributes

To address our second research question aimed at characterizing the quality and attributes of graphs constructed by participants, we categorized the graphs qualitatively based on similarities and differences that emerged across participants and participant groups. Since the purpose of a graph is to communicate information, graphs constructed by the participants during the think-aloud interview were examined as stand-alone artifacts (Figure 2.6 and Appendix 2D). The attributes used to evaluate graphs fell into four main categories: graph mechanics, data form, graph choice, and aesthetics and visuo-spatial considerations. Each category is described in more detail in Appendix 2C.

Graphs constructed by undergraduate students (UGR and UGNR) and graduate students (GS), but not professors, followed basic graph conventions, and included meticulously-labeled axes, titles, tick marks, scale, and key. Ten of the fifteen undergraduate students titled their graph whereas only one of the eight GS and one of the five professors titled their graph. In terms of axes labels, all participants labelled 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 labelling the axis, 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. UGNR4 and UGR3 did not plan ahead of time the space they were using or the scale because midway in the scaling process they realized that they were running out of space, so they decided to add an axis break. In contrast to the undergraduate and graduate students, professors tended to sketch their graphs, omitting detailed axis labels and meticulous plotting (Appendix 2D).

The graphs constructed by all participants were, in general, aesthetically sound and the presence of Gestalt principles 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 points labels (UGNR3, UGNR5, UGNR10, UGR1, GS8), which made it difficult to understand the take-home message. Graph constructed by UGNR3 was constructed in a manner that hindered communication such that it was difficult to identify the data points from the graph and formulate an accurate take home message (Appendix 2D).

Participants were given a table of data to construct a graph. This table consisted of one independent variable, one dependent variable, two treatments, and three replicates in each treatment (Appendix 2A). It is noteworthy that all undergraduate students except for UGR9138 plotted either some (UGNR2, UGNR7, UGNR8, UGR3) or all (UGNR4, UGNR1, UGNR3, UGNR5, UGNR 9, UGNR10, UGR1, UGR2), of the raw data values. In contrast, graduate students and professors (except for GS8 and P3) and UGR4, UGR5, and UGNR6 collapsed the data, plotting 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.

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*. We see that the graph created by UGNR2 did not align with the research question or hypothesis because they only plotted one plant at 15 ml and ignored the three plants in the treatment with 5 ml of water. Likewise, UGR3, UGNR7, and UGNR8 plotted the growth of bacteria in three tubes at only 22 degrees Celsius, entirely ignoring the data for the other three tubes at 10 degrees Celsius. Similarly, professor 3 only plotted tube 1 at 22 degrees Celsius, ignoring the other tubes and other treatment. All graphs constructed by graduate students aligned with the research question posed in the task.

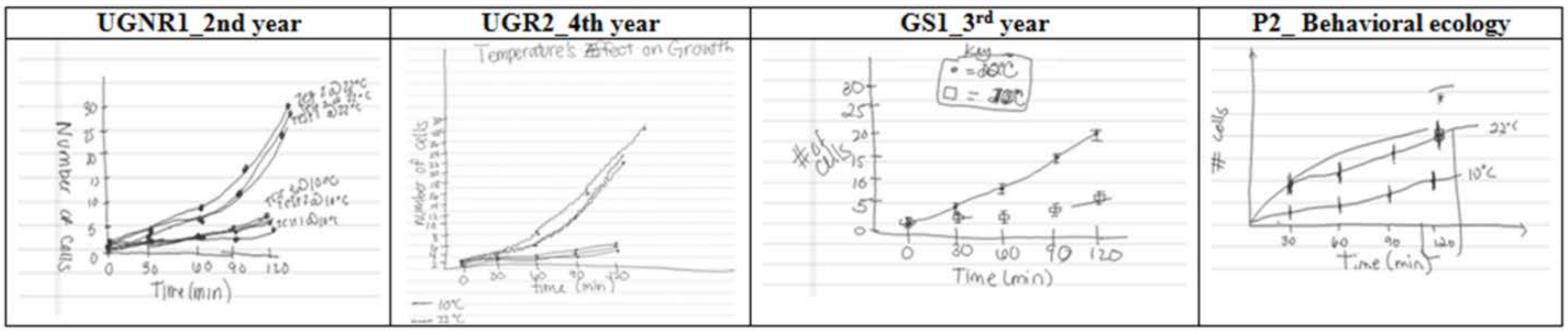


Figure 2.5: Graph Exemplars from all participant groups using the Bacteria Scenario: Examples of the types of graphs constructed by participants when given either the bacteria scenario or plant scenario. (A complete summary of the graphs constructed can be found in Appendix D).

2.8 Discussion

Recommendations for undergraduate biology education include engaging students in the practices of science as a critical feature of the learning of their discipline (Brewer and Smith 2009). Numerous published studies report the benefits of research experiences on student learning and engagement with science (Auchincloss et al., 2014; Lopatto, 2004; Lopatto and Tobias, 2010; Seymour et al., 2004). One of the science practices that is fundamental for science and conducting research is working with data, which includes understanding, consolidating and communicating data in graphic representations. Past studies have shown student difficulties with different aspects of graph choice and construction at the K-12 level (Leonard and Patterson, 2004; Tairab and Al-Naqbi, 2004; Padilla, McKenzie and Shaw, 1984; Brassell and Rowe, 1993; Ainley, 2000). Recent studies have corroborated these findings and added that these difficulties are present also in undergraduate biology students (McFarland, 2010; Bray-Speth et al., 2010; Angra and Gardner, 2014), medical doctors (Schriger and Cooper, 2001; Cooper, Schriger and Tashman, 2001; Cooper, Schriger and Close, 2002), and professionals (Rougier et al., 2014; Weissgerber et al., 2015). Although many instructional books exist on graphing (e.g. Tufte, 1983; Kosslyn, 1994; Few, 2004), they are not targeted towards teaching or improving the complex reasoning required for choosing and constructing proper graphs in the sciences. Further, the exact difficulties that people have in reasoning with data and graphs are not fully understood. In order to fill the graph construction gap, we aimed to understand the reasoning associated with graph choice and construction. In this study we used the Meta-Representation Competence (MRC) and expert-novice frameworks to understand how undergraduate students without research experience, with research experience, 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. Our findings suggest that being able to thoughtfully plan, think about data, and reason about the function of variables and to articulate the purpose of a graph, results in well-conceived graphs that communicate clear messages. Additionally, we found graphs that displayed the most processed data, i.e. averages with error bars, were usually created by participants who had research experiences, such as the graduate students and expert professors. Here we elaborate on our main findings in the context of MRC to illustrate the graph construction reasoning and the degree of graphing competence with a discussion of the implications for classroom instruction and student learning.

2.9 Patterns of MRC Categories Across Novices and Experts

2.9.1 Reflection

According to the MRC framework, when a participant reflects, they are demonstrating understanding of the data and representations at hand. This MRC component emerged in novices and experts across all three phases of the graph construction task (Figures 2.2-2.4 and Figure 2.5). In the planning phase, before graph construction took place, we see evidence of the MRC component, *reflection* as represented by the theme *data table*. In our study when novices and experts were referring to the *data table*, they were making sense of the data provided in the data table by summarizing the data and or the variables present (see Table 2.1 for quotes). Figures 2.2 and 2.5 indicate that the theme, *data table* was present across multiple individuals, suggesting making sense of the data is an important feature of the planning phase for our participants. This aligns with the first two steps in Polya's problem solving model (1945), understanding the problem and devising a plan and with Koedinger and Anderson's work (1990) on planning before attempting to solve proofs in geometry. This theme re-appeared in three undergraduates and GS1 during the reflection phase of the interview, when participants were asked to explain their reasoning behind graph choice (Figure 2.4). Quotations in Table 2.3 indicate that UGNR did not focus on the *function* of the graph but instead, used the *data table* as a way to link their reasoning to the graph they created. This observation is consistent with what diSessa (2004) calls the "quick and unreflective" instincts employed by students regarding representations, detail needed, and the message that needs to be communicated. This could be because in a classroom setting, students are not asked to reflect explicitly nor are they taught how to reflect. Without knowing how to recall and what relevant information is, the most novice students who lack both research and experimental experiences resort to form their explanations from the most recent information exposure, which in this case, is the data table.

In the construction phase, we saw new themes that align with the MRC component of *reflection* emerge. These are: *evaluation* and *technology*, defined in Table 2.2. Figures 2.3 and 2.5 show multiple instances of *evaluation* in expert professors in the construction phase and multiple instances from all four populations in the reflection phase. When participants were evaluating, they were either reflecting on their general graphing habits, future directions, or the take home message. Although all participants revealed their level of expertise with experimental design, difficulties pertaining to *variables* and understanding the *purpose* of the graph emerged for UGNR. This is evident in Figure 2.4, where the themes *variables* and *purpose* are not apparent during the reflection phase.

While it is the norm to perform data analysis and create graphs using computers, two undergraduate students with research experience and one graduate student noted this, as indicated by the theme *technology* in the construction phase (Figures 2.3 and 2.5). This could be because in their research lab and in the classroom setting where they are asked to create graphs or data representations, they use software programs such as Microsoft Excel. Professors surprisingly did not mention technology in the same way, rather at one point in the interview, Professor 4 said *"I wish my students were here watching me do this because they think you have to have a software package to even think about data."* This observation highlights the potential tension between technology and undergraduate education as it pertains to the practices of scientists. Policy documents encourage instructors to utilize and teach with technology (NRC 2003) because it engages students with the content material and prepares them for the real world. Although technology is ideal when students are learning other tasks like mathematical modeling, it may be detrimental to introduce it to novice learners as the first step, or the only approach in graphing. As Professor 4 and previous literature (Deacon 1999; Leonard and Patterson 2004) suggest, it may be better to first plan and sketch a graph by hand to understand the data before diving in with graph construction using software.

2.9.2 Function

The MRC component *function* refers to having an extensive knowledge of different types of representations, their usage, advantages, and disadvantages. A common theme that occurred across all three participant groups and phases was *graph choice* (Figure 2.5). In the planning phase, this theme was present only for two undergraduate students with research experience (UGR1 and 3), UGNR4, and GS4 and 6. Immediately after reading the scenario, all five participants decided on a graph type before proceeding. Professors did not immediately decide on a graph type, but first decided to articulate the research question and the *purpose* of the graph. In the construction phase, we saw *graph choice* emerge in GS 1, 4, 7, and 8, UGR2 and 3, UGNR 7 and 10. We did not see the category of *graph choice* emerge from the professor group because the professors did not explicitly state their graph choice, but implicitly had a graph type in mind as they plotted data. Since in the last phase of the interview, participants were asked to reflect on graph choice, it was present in all participants. Another theme under the category *function* was *time*. Almost all participants with the exception of GS2, GS3, GS4, GS5, GS7, UGNR2. UGNR6, UGR4, and UGR5 created a line graph (Appendix 2D). All participants, except for UGNR3, correctly identified time as the independent variable and labelled the x-axis. When asked to articulate the reasoning behind their graph choice, professors and UGR spoke in terms of the variable time and explained that a line graph allows them to "*see a change in or compare the growth over time*" (Table 2.2). This is true of line graphs that are useful for seeing a change as a function of independent variables such as time and temperature (Franzblau and Chung, 2012). Finally, in their reflection, professors and graduate students talked about the *purpose* of the graph they were constructing. The UGNR did not justify their reasoning behind graph choice with the *variables* displayed on the graph, but rather with their intuition (see Table 2.2).

2.9.2 Invention

The MRC component *invention* consists of creating a new representation but with conceptual and or cultural rules (diSessa and Sherin, 2000). In this study, we asked our participants to translate a table of numbers that came from an experiment designed to answer a question into a graph of their choice. In order to create a meaningful graph, understanding the function of the graph as it relates to the variables, idea of the type of data to display, and appropriate mechanics are necessary to represent the data in the graph best. All participants utilized their prior knowledge or constructive resources (Sherin, 2000; Azevedo, 2000) to shape how they designed the graph.

Most participants created a line graph (see Figure 2.6 and Appendix 2D). Previously Padilla, McKenzie, and Shaw (1986) stated that skills required for constructing line graphs were:

1. Drawing and scaling axes;
2. Assigning variables to correct axes;
3. Plotting points;
4. Using a line of best fit.

Addressing the graphing skills mentioned by Padilla, McKenzie, and Shaw (1986), we found that meticulously constructed and scaled axes were present in all of the undergraduate and graduate students' graphs, but not present in the graphs constructed by the professors. This could be because unlike the undergraduate students who spent time plotting all of the raw data points, data plotted by professors were summaries of transformed data (Figure 2.6 and Appendix 2D).

Although graph mechanics are necessary components that guide the graph interpreter, their addition to the graph generally does not require deep reflective consideration as compared to other decisions that go into graph construction. Graphs constructed by graduate students consisted of elements from both the undergraduate and professor groups. Similar to the undergraduate students, graduate students' graphs consisted of meticulous graph mechanics such as scaled axes. However, the transformed data plotted focused on the big-picture trends and resembled the data choice used by expert professors. We also found that almost all graphs constructed in this task lacked a descriptive title. This is not necessarily surprising given the informal nature of our task and that graphs in textbooks and journal articles do not include titles above the graph, but rather as the first sentence of the figure legend (Rybarczyk, 2011).

Assigning variables to correct axes was not an issue for many participants, except for UGNR3 who went back and forth when deciding what variable to plot on which axes. When probed to explain the labeling of the graph, UGNR3 exclaimed that *"because we are plotting leaves vs. time, which I didn't realize at first…"* tells us that a possible reason that they didn't realize this at first was because they failed to understand the problem and devise a plan before trying to solve the problem, which affected the aesthetic quality and accuracy of the data and the take home message. Although only one participant in our study displayed difficulty with variables, it is still an important finding that shows that graphing difficulties documented in middle school (Padilla, McKenzie and Shaw, 1986) can still persist well into higher education.

Another important decision to consider when inventing a graph is the purpose of the graph, which dictates the type of data that will be displayed in the graph and consequently the type of statistics (i.e. descriptive or inferential) and conclusions that can be inferred from the graph (Figure 2.5). In the interview scenario, participants were presented with a dataset that contained one independent variable, one dependent variable, two treatments, with three replicates in each treatment and asked to plot a graph showing either the effect of temperature on bacteria growth or the effect of water on plant growth. Previous research at the K-12 level in statistics education reveals that students perceive graphs as a “collection of points” and focus on individual data rather than generalizing the trends of the entire dataset (Cobb, 1999). Interestingly, in our study, we found that all UGNR and UGR except for UGNR2, UGNR6, UGR4, and UGR5 viewed the data in the data table as separate entities or case values (Konold et al., 2015) and plotted all of the raw data presented in the data table. As Konold et al. (2015) reveal in their study, perceiving data as case value is a naïve way of examining data and is evident in young children who stress the importance of each individual data point. While viewing data as case value can be useful because it keeps the student connected to the data (Lehrer and Schauble, 2004), it can cause the student to lose meaning from the entire spread of data. This was also evident when participants were asked to articulate their reasoning for data plotted on the graph. Like the other participants, UGR2 said, “*I just plotted the number of cells um at each point in time that was given to me.*” UGNR2 plotted one plant that received 15 ml of water per day in a scatter plot (Appendix 2D). When asked to articulate reasoning for data plotted, the answer was succinct and vague, despite the interviewer probing for elaboration: “*I was plotting the number of leaves vs the time or the hours in a day.*” Even though all participants were not given explicit instructions on which data to plot or in what form, it is interesting to note that most students immediately plotted all of the raw data without considering data transformation, but the ones who did manipulate the data reflected on statistics (Figure 2.5). One possibility for this behavior could be because participants were not given explicit instruction. In the classroom (Bray-Speth et al., 2010) and even on standardized exams (College Board AP Biology, 2016), students are always given explicit directions on the expectations of the task. Moreover, students are well equipped with technology in a classroom setting to transform data quickly, if instructed to do so. Even though we provided our participants with a simple dataset, even if a preliminary, internal thought of transforming data had occurred, doing so without a calculator may have intimidated some participants and caused them to resort to plotting raw data.

Conversely, UGR4, UGR5, and UGNR6 took the time to calculate averages and plot these transformed data values into the graph (Appendix 2D). When asked to explain the data plotted in the graph, UGR4responded: "S*o each of these points is the average of the number of leaves of the three plants at that water level and that time. So first I took the average of the number of leaves of the three plants because that's like three trials, so then you can have like a sort of like a general idea of like how many leaves are gonna be made for each so then you can plot that like general idea*." From this quotation, it is evident that UGR4 wanted to generalize the trends in order to answer the original prompt aimed to understand "how the amount of water influences plant growth". Although the scatterplot that was created with the average values did not have error bars, UGR4 did sketch a trendline in order to generalize the trends, which shows their knowledge of inferential statistics and extrapolating to a larger population of plants.

Professors and graduate students like UGR4, UGR 5, and UGNR6 plotted transformed data, except for P3 GS4, and GS8. Professor 3clearly stated that they were interested in understanding how tube 1 grew at 22 degrees Celsius (Appendix 2D).

Unlike UGNR6, UGR4 and GS2, professors and other graduate students who displayed the averages and accompanied their data points with error bars or made explicit that if they had time or were creating this graph for a formal presentation, that they would add the error bars. The only undergraduate student who accompanied their averaged data with error bars was UGR5. By transforming the data and plotting the averages or summations, professors and graduate students as a whole viewed the data with a different lens than the undergraduate students (Figure 2.5; Konold and Higgins, 2003; Konold et al., 2015). When asked why averages were plotted, Professor 5 said, *"most of the time it's the simplest or most obvious thing to do and it's not necessarily the best thing to do depending on what the data set looks like here we only have 3 numbers and the numbers are relatively similar."*

By collapsing the replicates across each treatment, Professor 5 as well as other professors and graduate students implicitly indicated that they understood the purpose of the replicate is to show multiple experimental runs under each experimental treatment. By plotting the transformed values, professors maintained their focus on the purpose of the task and focused on the overall trend seen from the two treatment groups rather than from the individual plant or bacteria. Professor 2 did not calculate the averages from the data table or scale their y-axis. Instead, they segmented the data in the table into two separate treatments and used the values provided as a guideline. They sketched two lines, error bars and mentioned that they would run an inferential test to see if there was an actual difference between the two treatments. Although the professor did not spend time calculating averages, their reasoning behind sketching averages resonated with that of the other professors in the study. Graphs created by professors answered the prompt more so than the undergraduate student's graphs (Figure 2.6). However, as uttered by Professor 5, immediately deciding on averages as the "*most obvious thing to do*", may not be quite so obvious to the novice learner, or in this case, the UGNR, who lack experiences with data and transforming data. This finding adds support to the need for providing undergraduate students with multiple experiences with data.

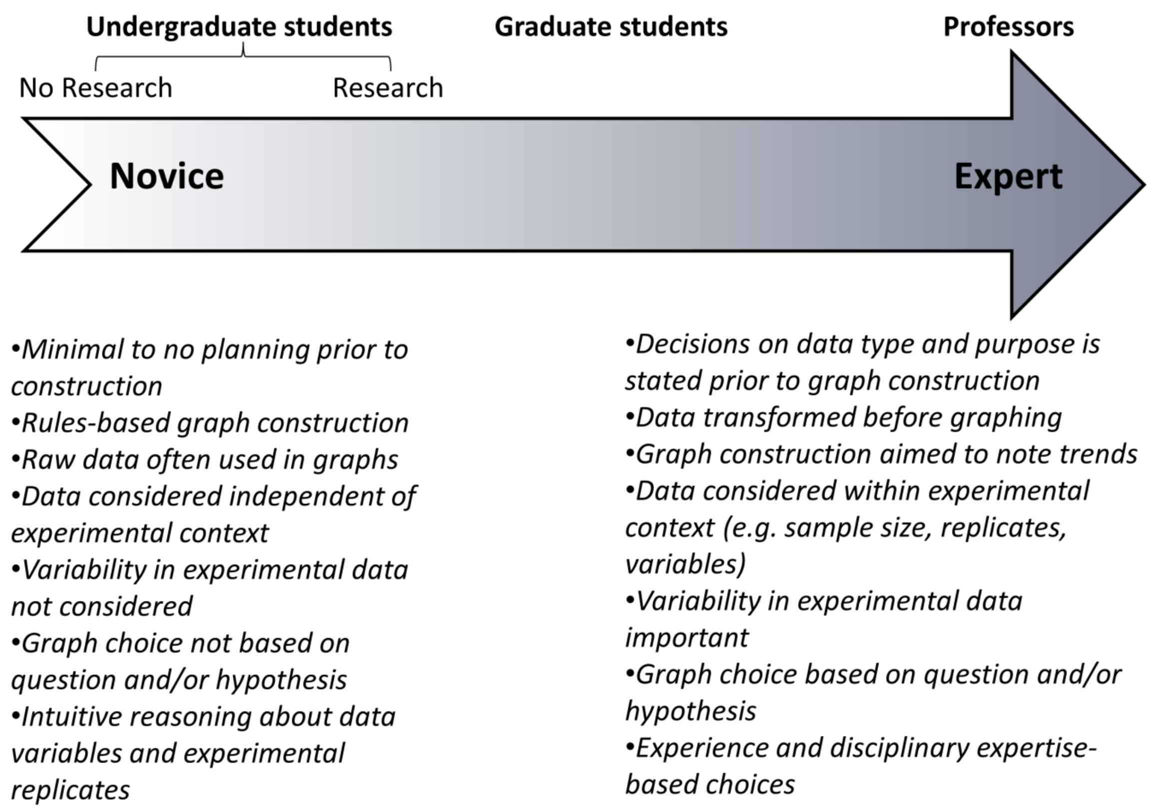


Figure 2.6: Summary of Graph Construction Reasoning and Graph Attributes: Visual summary of findings with the reasoning behind graph choice and construction along the novice to expert continuum.

2.9.3 Critique

The MRC component critique is the critical knowledge that is essential for assessing the quality of representations (diSessa, 2004). In a past study by diSessa (2004), it was reported that when students critique representations, they focus on two things: parsimony and completeness of a representation. Although in the present study we did not ask participants to explicitly critique their representation, we noticed this category during the construction phase with the themes of *sample size* and *aesthetics* (See Figures 3 and 5). Specifically, professors critiqued the small sample size in the data table. This reasoning most likely stems from the professor's content knowledge that reflects a deep understanding of experimental design and statistics. The undergraduate students and GS8 directed their critique on visual aspects of their graphs, focusing on the *aesthetics* (Tufte, 2001).

2.10 Summary of MRC in Graphing

Implicit in the MRC framework is expert competence with creating and understanding representations. While all participants were able to engage in reasoning within all MRC categories, there is evidence for expert-novice differences across our participant groups (Figures 5 and 7). All professors took time to understand the data before proceeding with graph construction, 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 graph, 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 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 done in the context of neurobiology which 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 Bransford et al. (2000), who state that experts organize their knowledge that reflects a deep understanding of the subject matter and expert knowledge cannot be recalled as a set of isolated facts but it 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 in the undergraduate students. Likewise, in Hay et al. (2013) study, neuron drawings by the laboratory leaders were original and unlike that of ones found in textbooks because they were informed by years of experiences and accumulated knowledge.

2.11 Project Scope and Future Studies

Four main study design features bound the scope of our conclusions. First, data were collected from expert and novice biologists at a single Midwestern United States research university, which is a unique environment with its own curriculum and student population. Furthermore, since our study consisted of a small group of participants, the claims we presented are not broad generalizations to the types of things that all experts or novices do or think. We are stating our findings as illustrated by ten undergraduate students with no research experience (UGNR), five with research experience (UGR), eight graduate students (GS), and five professors. 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 in order to fully understand and appreciate what the reasoning is like for graph choice and construction.

Second, for the purpose of our study, we provided all participants with a simple dataset with one independent variable, one dependent variable, two treatments, with three replicates each. In order to replicate our study but 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 UGR 3 showed difficulty with scaling axes (Padilla, McKenzie and Shaw, 1984; Li and Shen, 1992; Brassell and Rowe, 1993; Ainley, 2000) as indicated by the awkward positioning of the axis breaks, and UGNR3 showed difficulty with variables, as indicated by their graph (Tairab and Al-Naqbi, 2004; Appendix 2D).

However, the simplicity of the dataset may have caused Professor 2 to go into “teacher mode” and quickly sketch the data in order to illustrate how temperature influences bacteria growth, instead of taking their time to plot data.

Third, participants in our study were given a dataset. 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 dataset and with data the participants collected themselves.

Lastly, participants in this study constructed graphs manually using a LiveScribe pen and paper, instead of using the modern and conventional method of graph construction on the computer. Having participants narrate their thought process during manual construction allowed us to understand their reasoning fully. If we had asked participants to construct graphs using software programs, it may have tampered with their graph choice by biasing them towards graph choices presented by the software package. By using manual construction, we were able 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, Professor 2 in particular, who studied the data table, then sketched the data with error bars in order 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 reveal areas of difficulties and competencies with graphing.

2.12 Conclusions and 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 above. Further, the graphs produced by participants in the study also varied along the novice-expert continuum. Figure 2.7 summarizes the attributes of the graph construction reasoning and the graphs that we observed in the most novice and most expert participants. The distinctions summarized in this figure highlight potential target areas for instructors to promote more expert-like data handling and graphing practices.

As more undergraduate students are encouraged to engage in inquiry-based biology labs and seek research opportunities during their higher education, 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 towards expertise. In order for students to feel comfortable and take advantage of first-year research inspired labs, graphing practices especially focusing on the MRC categories should be implemented to foster the development of expertise. Although undergraduate research experiences differ widely (Lopatto, 2004), our findings suggest that having these experiences may positively impact students' decisions with data and graphing. Since not all institutions offer these types of experiences, we encourage instructors to find creative ways to incorporate data into their lessons. One way to promote MRC is to train instructors at the K-12 and undergraduate levels to articulate their reasoning multiple times as they are teaching students about graphs. Providing students with ample opportunities to invent, critique and reflect on the functions of graphs with targeted feedback is also critical for developing expertise. Although two attempts have been documented with biology undergraduate students in lecture (Bray-Speth et al., 2010) and laboratory (McFarland, 2010), these studies did not utilize MRC-focused activities and instruction to instill critical reflective practices during graphing. Furthermore, instructional tools have been published in the literature but they lack the reflective component of the MRC (Puhan et al., 2006; Paniello et al., 2011; Webber et al., 2014; Create A Graph Tutorial-NCES Kids' Zone, 2016; Interactive Statistics Map, 2016). In a previous article (Angra and Gardner, 2016) we provide novel instructional and learning tools that offer students and instructors guidance towards appropriate graphing practices. The inspiration behind these tools were the MRC framework (diSessa and Sherin, 2000; diSessa, 2004) and Polya's problem solving model (Polya, 1945), with the ultimate goal of moving students towards expertise. Using these tools to emphasize graph choice and construction skills will encourage students to think critically about data and graphs in and outside of 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 towards 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).

Acknowledgements

We would like to thank Dr. Kathryn Obenchain for her qualitative research expertise and constructive feedback on early drafts of this manuscript. We thank Ms. Janetta Greenwood for helping us decide on the plant and bacteria scenarios. We are indebted to all of the biology undergraduate students and professors who participated in this study. We also thank our research group, PIBERG, for their feedback on this project. This project emerged from ideas initiated within the Biology Scholars Research Residency program (SMG).

References

AAMC-HHMI Committee. (2009). Scientific foundations for future physicians. Washington, DC: Association of American Medical Colleges.

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. Mahwah, NJ: Lawrence Erlbaum Associates.

Ainley, J., Nardi, E., & Pratt, D. (2000). The construction of meanings for trend in active graphing. International Journal of Computers for Mathematical Learning, 5(2), 85-114.

Ainley, J. (2000). Transparency in graphs and graphing tasks: An iterative design process. The Journal of Mathematical Behavior, 19(3), 365-384.

Aliaga et al. (2005). Guidelines for Assessment and Instruction in Statistics Education College Report. American Statistical Association.

Angra A, Gardner SM. (2014). Using Student Laboratory Reflections to Reveal Reasoning Behind Graph Choice and Construction. West Lafayette, IN: Purdue University Chapter. The Society of Sigma Xi, Poster.

Angra, A., & Gardner, S.M. (2016) The Development of a Framework for Graph Choice and Construction. Advances in Physiology Education.

Auchincloss, L. C., Laursen, S. L., Branchaw, J. L., Eagan, K., Graham, M., Hanauer, D. I., ... & Towns, M. (2014). Assessment of course-based undergraduate research experiences: a meeting report. CBE-Life Sciences Education, 13(1), 29-40.

Azevedo, F. S. (2000). Designing representations of terrain: A study in meta-representational competence. The Journal of Mathematical Behavior, 19(4), 443-480.

Bangera, G., & Brownell, S. E. (2014). Course-Based Undergraduate Research Experiences Can Make Scientific Research More Inclusive. CBE-Life Sciences Education, 13(4), 602-606.

Beichner, R. J. (1994). Testing student interpretation of kinematics graphs. American journal of Physics, 62(8), 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(6), 527-554.

Bertin, J. (1983). Semiology of graphics: diagrams, networks, maps. Madison, WI: Univ. of Wisconsin Press.

Bowen, G. M., Roth, W. M. Data and graph interpretation practices among preservice science teachers. (2005). Journal of research in science teaching 42, 1063-1088.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). How people learn. Washington, D.C.: National Academy Press.

Brasell, H. M., & Rowe, M. B. (1993). Graphing skills among high school physics students. School Science and Mathematics, 93(2), 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.

Brewer, C. A., & Smith, D. (2011). Vision and change in undergraduate biology education: a call to action. American Association for the Advancement of Science, Washington, DC.

Bright, G. W., & Friel, S. N. (1998). Graphical representations: Helping students interpret data. Reflections on statistics: Learning, teaching, and assessment in grades K–12, 63-88.

Brownell, S. E., Hekmat-Scafe, D. S., Singla, V., Seawell, P. C., Imam, J. F. C., Eddy, S. L., Stearns,T. & 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(2), 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(4), 473-493.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. Cognitive psychology, 4(1), 55-81.

Chi, M. T. (2006). Two approaches to the study of experts' characteristics. The Cambridge handbook of expertise and expert performance, 21-30.

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(5), 290-295.

Cleveland, W. S. (1984). Graphical methods for data presentation: Full-scale breaks, dot charts, and multibased logging. The American Statistician, 38(4), 270-280.

Cobb, P. (1999). Individual and collective mathematical development: The case of statistical data analysis. Mathematical thinking and learning, 1(1), 5-43.

College Board. AP Biology. Course and Exam Description (online). Retrieved August 05, 2016, from https://secure-media.collegeboard.org/digitalServices/pdf/ap/ap-biology-course-and-exam-description.pdf

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(3), 259–267. http://doi.org/10.1187/cbe.10-11-0137

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.

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.

Create A Graph Tutorial-NCES Kids' Zone. (n.d.). Retrieved August 05, 2016, from http://nces.ed.gov/nceskids/help/user_guide/graph/index.asp

Creswell, J. W. (2013). Qualitative inquiry and research design: Choosing among five approaches. Sage publications, Inc.

Deacon, C. (1999). The importance of graphs in undergraduate physics. The physics Teacher, 37(5), 270-274.

diSessa, A. A. and Sherin, B. L. (2000). Meta-representation: An introduction. The Journal of Mathematical Behavior, 19(4), 385-398.

diSessa, A. A. (2004). Metarepresentation: Native competence and targets for instruction. Cognition and instruction, 22(3), 293-331.

Ericsson, K. A., & Smith, J. (1991). Toward a general theory of expertise: Prospects and limits. Cambridge University Press.

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

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. Psychological review, 100(3), 363.

Ericsson, K. A. (2006). The influence of experience and deliberate practice on the development of superior expert performance. The Cambridge handbook of expertise and expert performance, 38, 685-705.

Few, S. (2004). Show me the numbers. Oakland, CA: Analytics Press.

Friel, S. N., & Bright, G. W. (1996). Building a Theory of Graphicacy: How Do Students Read Graphs?

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. The Journal of hand surgery, 37(3), 591-596.

Garcia-Mila, M., Marti, E., Gilabert, S., & Castells, M. (2014). Fifth Through Eighth Grade Students' Difficulties in Constructing Bar Graphs: Data Organization, Data Aggregation, and Integration of a Second Variable. Mathematical Thinking and Learning, 16(3), 201-233.

Grawemeyer, B., & Cox, R. (2004, March). The effect of knowledge-of-external-representations upon performance and representational choice in a database query task. In International Conference on Theory and Application of Diagrams (pp. 351-354). Springer Berlin Heidelberg.

Green, J. L., Camilli, G., & Elmore, P. B. (2012). Handbook of complementary methods in education research. Routledge.

Grunwald, S., & Hartman, A. (2009). A case-based approach improves science students' experimental variable identification skills. Journal of College Science Teaching, 39(3), 28.

Hatch, J. A. (2002). Doing qualitative research in education settings. 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(3), 468-491.

Hoffman, K., Leupen, S., Dowell, K., Kephart, K., & Leips, J. (2016). Development and Assessment of Modules to Integrate Quantitative Skills in Introductory Biology Courses. CBE-Life Sciences Education, 15(2), ar14.

Humphrey, P. B., Taylor, S., & Mittag, K. C. (2014). Developing consistency in the terminology and display of bar graphs and histograms. Teaching Statistics, 36(3), 70-75.

IBM Corp. Released. (2013). IBM SPSS Statistics for Windows, Version 22.0. Armonk, NY: IBM Corp.

Interactive Statistics Map. (n.d.). Retrieved August 05, 2016, from http://bioquest.org/numberscount/statistics-map-2/

Kanari, Z., & Millar, R. (2004) Reasoning from data: How students collect and interpret data in science investigations. Journal of Research in Science Teaching, 41(7), 748–769.

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

Koedinger, K. R., & Anderson, J. R. (1990). Abstract planning and perceptual chunks: Elements of expertise in geometry. Cognitive Science, 14(4), 511-550.

Kosslyn, S. M. Elements of graph design. New York: WH Freeman, 1994.

Konold, C., & Higgins, T. L. (2003). Reasoning about data. A research companion to principles and standards for school mathematics, 193-215.

Konold, C., Higgins, T., Russell, S. J., & Khalil, K. (2015). Data seen through different lenses. Educational Studies in Mathematics, 88(3), 305-325.

Kostelnick, C. (1998). Conflicting standards for designing data displays: Following, flouting, and reconciling them. Technical communication, 45(4), 473-473.

Lehrer, R., & Schauble, L. (2004). Modeling natural variation through distribution. American Educational Research Journal, 41(3), 635-679.

Leinhardt, G., Zaslavsky, O., & Stein, M. K. (1990). Functions, graphs, and graphing: Tasks, learning, and teaching. Review of educational research, 60(1), 1-64.

Leonard, J. G., & Patterson, T. F. (2004). Simple computer graphing assignment becomes a lesson in critical thinking. NACTA journal, 48(2), 17-21.

Li, K. Y., & Shen, S. M. (1992). Students' weaknesses in statistical projects. Teaching Statistics, 14(1), 2-8.

LiveScribe, Inc. (2015) LiveScribe smartpens. Retrieved from http://www.livescribe.com/en-us/smartpen/

Linn, M. C., Palmer, E., Baranger, A., Gerard, E., & Stone, E. (2015). Undergraduate research experiences: Impacts and opportunities. Science, 347(6222), 1261757.

Lopatto, D. (2007). Undergraduate research experiences support science career decisions and active learning. CBE-Life Sciences Education, 6(4), 297-306.

Lopatto, D. (2004). What undergraduate research can tell us about research on learning. In What Works, What Matters, and What Lasts 4, 1-7. Retrieved December 9, 2005, from http://www. pkal. org/documents/Vol4WhatUndergradResearchCanTellU s. cfm Reynolds, C., III.

Lopatto, D., & Tobias, S. (2010). Science in solution: The impact of undergraduate research on student learning. Council on Undergraduate Research.

Lyons, N. (2010). Handbook of Reflection and Reflective Inquiry.

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(3), 317–326. http://doi.org/10.1187/cbe.07-07-0046

National Institutes of Health Research Training and Career Development Retrieved August 05, 2016, from https://researchtraining.nih.gov/programs/training-grants

National Science Foundation Research Traineeship Program (NRT) Retrieved August 05, 2016, from https://www.nsf.gov/funding/pgm_summ.jsp?pims_id=505015

National Research Council (US) Committee on Undergraduate Biology Education to Prepare Research Scientists for the 21st Century. (2003).BIO2010: Transforming undergraduate education for future research biologists. National Academies Press (US).

Nuzzo, R. L. (2016). The Box Plots Alternative for Visualizing Quantitative Data. PM&R, 8(3), 268-272.

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(1), 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(6), 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(1), 95-115.

Patton, M. Q. (2005). Qualitative research. John Wiley & Sons, Ltd.

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 model.

Preece, J., & Janvier, C. (1992). A study of the interpretation of trends in multiple curve graphs of ecological situations. School Science and Mathematics, 92(6), 299-306.

Puhan, M. A., ter Riet, G., Eichler, K., Steurer, J., & Bachmann, L. M. (2006). More medical journals should inform their contributors about three key principles of graph construction. Journal of clinical epidemiology, 59(10), 1017-e1.

Roth, W. M., & McGinn, M. K. (1997). Graphing: Cognitive ability or practice?. Science Education, 81(1), 91-106.

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

Roth, W. M., & Bowen, G. M. (2003). When are graphs worth ten thousand words? An expert-expert study. Cognition and Instruction, 21(4), 429-473.

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

Rybarczyk, B. (2011). Visual literacy in biology: a comparison of visual representations in textbooks and journal articles. Journal of College Science Teaching, 41(1), 106-114.

Saxon, E. (2015). Beyond bar charts. BMC biology, 13(1), 60.

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., Shahi, D. & Osman, M. A. (2008). Transforming the teaching of science graduate students through reflection. Journal of College Science Teaching, 38(1), 32-36.

Seidman, I. (2013). Interviewing as qualitative research: A guide for researchers in education and the social sciences. Teachers college press.

Seymour E, Hunter A-B, Laursen SL, DeAntoni T. Establishing the benefits of research experiences for undergraduates in the sciences: first findings from a three year study. Sci Educ 2004; 88:493-534.

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(4), 690.

Sherin, B. L. (2000). Meta-representation: An introduction. The Journal of Mathematical Behavior, 19(4), 385-398.

Slutsky, D. J. (2014). The Effective Use of Graphs. Journal of wrist surgery, 3(02), 067-068.

Strauss, A., & Corbin, J. (1998). Basics of qualitative research: Techniques and procedures for developing grounded theory. Sage Publications, Inc.

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

Thiry, H., Weston, T. J., Laursen, S. L., & Hunter, A. B. (2012). The benefits of multi-year research experiences: differences in novice and experienced students' reported gains from undergraduate research. CBE-Life Sciences Education, 11(3), 260-272.

Tufte ER. (1983). Visual Display of Quantitative Information. Cheshire: Graphic Press.

Wainer, H. (2013). Medical Illuminations: Using Evidence, Visualization and Statistical Thinking to Improve Healthcare. Oxford University Press.

Webber, H., Nelson, S. J., Weatherbee, R., Zoellick, B., & Schauffler, M. (2014). The Graph Choice Chart. The Science Teacher, 81(8), 37.

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.

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 Biol, 14(4), e1002430.

Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. International Statistical Review, 67(3), 223-248.

Appendix 2A

Bacterial growth scenario

Imagine you are a microbiologist. You are particularly intrigued by how temperature affects the growth of bacteria. In order to answer your question, you set up an experiment that measures the growth of a particular type of bacteria at two different temperatures. You collect your data and display it in a chart shown below:

	Number of Cells					
Time (min)	22 °C			10°C		
	Tube 1	Tube 2	Tube 3	Tube 1	Tube 2	Tube 3
0	2	2	1	2	1	2
30	4	4	3	2	2	3
60	6	8	6	2	2	3
90	12	16	12	2	3	4
120	24	30	22	4	5	6

Plant leaves scenario

Imagine you are a botanist. You are particularly intrigued by how the amount of water influences plant growth. In order to answer your question, you set up an experiment that measures the growth of a particular type of plant at two different water amounts. You collect your data and display it in a chart shown below:

	Number of Leaves					
Time (Hours)	15 ml of water/day			5ml of water/day		
	Plant 1	Plant 2	Plant 3	Plant 1	Plant 2	Plant 3
0	0	1	0	0	0	0
30	2	1	3	0	1	1
60	3	3	5	2	1	2
90	4	3	5	2	1	3
120	6	5	7	3	1	4

Appendix 2B: Interview Protocol and Questions

Ice-breaker Questions:

1. Major, including sub-discipline:_______________
2. Have you taken (insert the pre-requisite classes required for this study)?
3. Year in college: _______________
4. Math classes: ____________________
5. Do you enjoy math classes in general? Why or Why not?
6. Current STEM classes:____________________________
7. Research experience:____________________________
8. Type of data most commonly handled: ______________________
9. Plans after graduation
10. Is there any other information you'd like to provide or update?
11. What is your first memory of learning how to create or use a graph? (grade school?)
12. How long have you been working with graphs?
13. What do you think graphs are used for?
14. Do you use graphs in your everyday life?
15. How do you construct graphs? For example, do you make graphs by hand, use computer programs, calculators, etc.

Task 1: Graph Construction

1. I am going to show you a data set with a brief description and then I would like you to create a graph from the data set. You can make a graph about any aspect of the data set that you would like. Take as much time as you want, don't rush, just relax. Be sure to talk through your thought process and what you are writing and drawing.

Show the participant one of two data sets, randomly chosen before the interview

As the participant works through the process, the interviewer will ask for clarifications and remind the participant to verbally articulate what they are thinking and what they are doing.

Questions to ask during the reflection phase:

a. Why did you decide to create the graph that you did?

b. What are you plotting? (raw data, computed value, etc.)

Appendix 2C: Categories used to evaluate graphs.

1. Graph mechanics
    a. Title- A title should be descriptive for the graph.
    b. Axes Labels- Both the x and y axes labels should be appropriate and descriptive for the experiment.
    c. Units- Should be appropriate and descriptive for the type of data displayed.
    d. Scale- Should be appropriate for the data displayed such that the increments are clear and easy to understand.

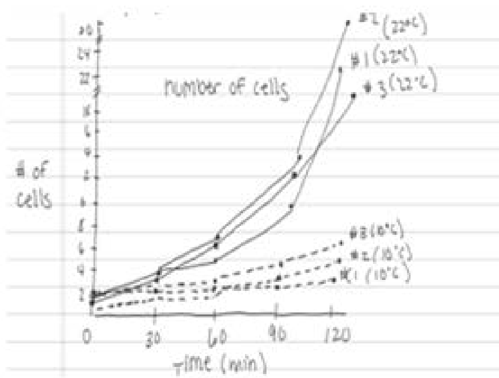
2. Data form
    a. Graph should show a clear distinction between raw and manipulated data plotted

3. Graph choice
    a. Graph Type- Graph type should be appropriate for both the independent and dependent variables.
    b. Alignment- Graph should align with the original intended purpose.
    c. Take-home message- graph type allows reader to draw appropriate conclusions from the data in the graph.

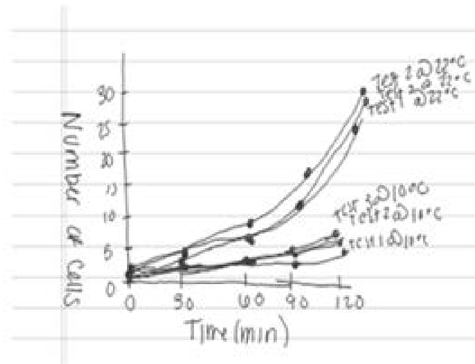
4. Aesthetics and Visuo-spatial Considerations
    a. The graph should be pleasing to the eye such that the data plotted occupy sufficient room in the Cartesian plane.
    b. Sound construction and mechanistic properties enable the reader to extract meaning from the graph

Appendix 2D: Graphs Constructed by Participants in Think-aloud Interviews
Graphs constructed from the bacteria scenario for undergraduates with no research experience (UGNR)

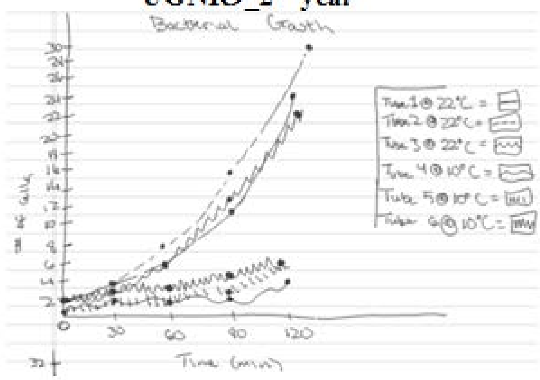
UGNR4_1st year



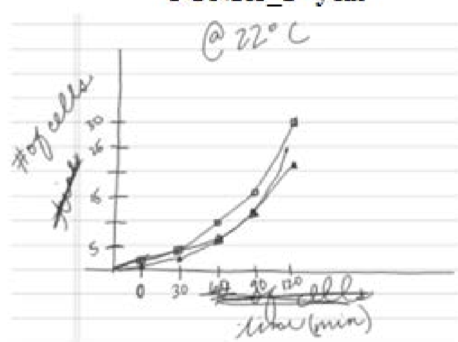
UGNR1_2nd year



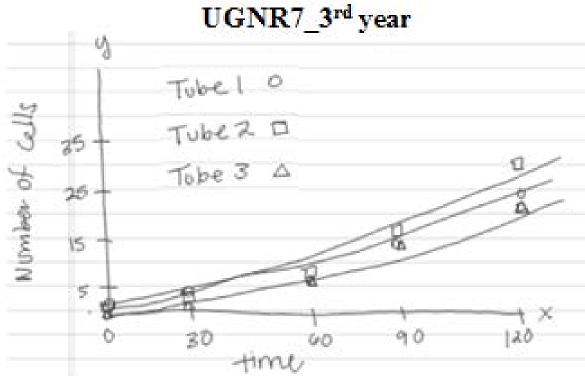
UGNR9_2nd year



UGNR8_1st year



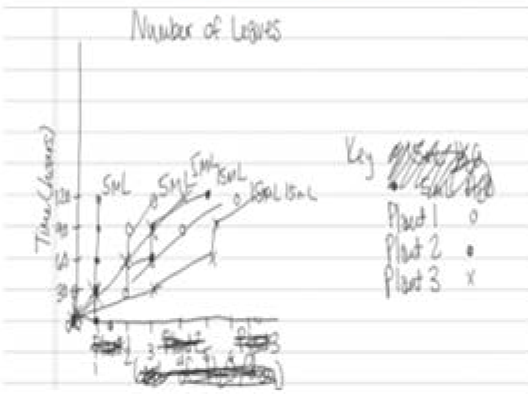
UGNR7_3rd year



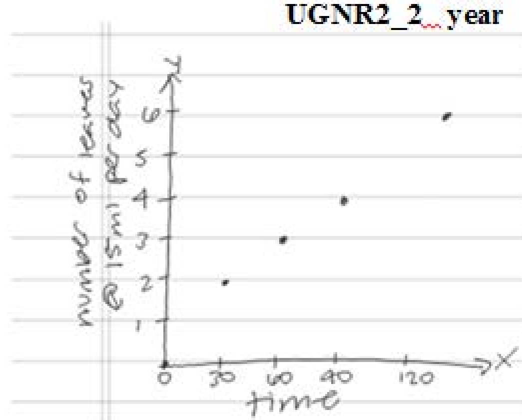
Appendix 2D Continued

Graphs constructed from the plant scenario for undergraduates with no research experience (UGNR)

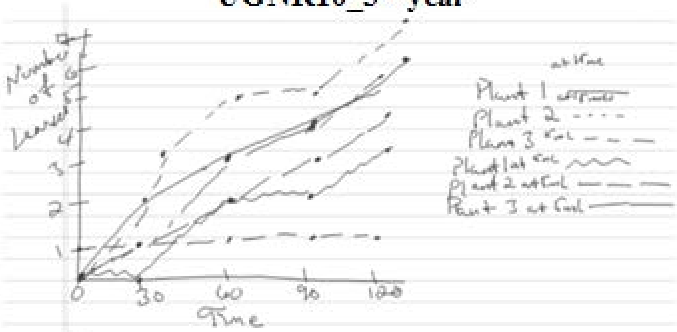
UGNR3_1st year



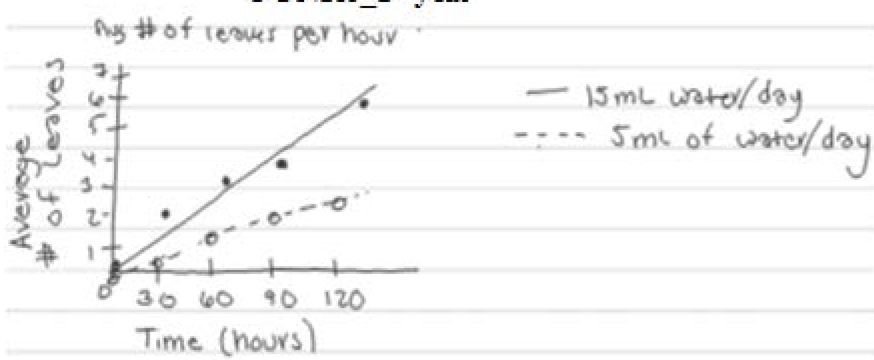
UGNR2_2nd year



UGNR10_3rd year



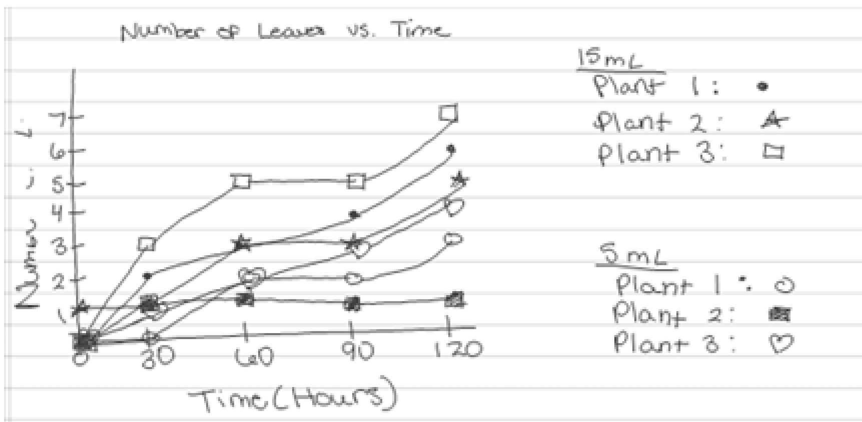
UGNR6_1st year



Appendix 2D Continued

Graphs constructed from the plant scenario for undergraduates with no research experience (UGNR)

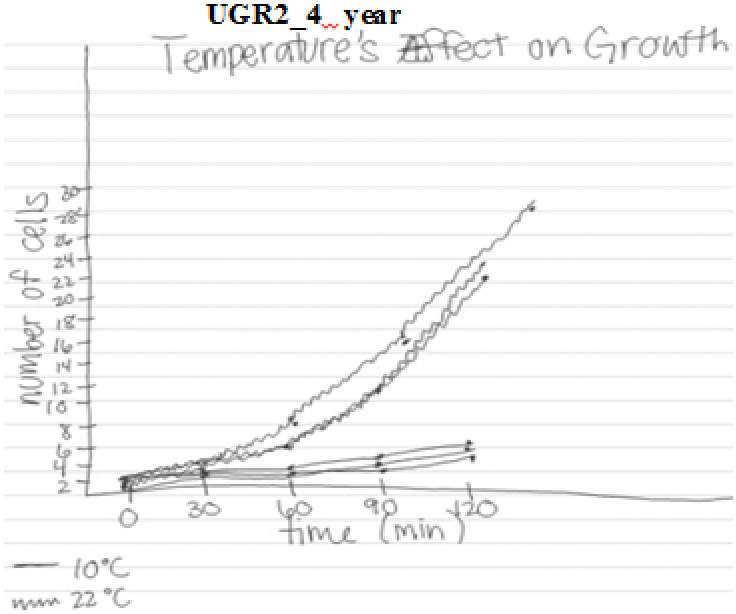
UGNR5_1st year



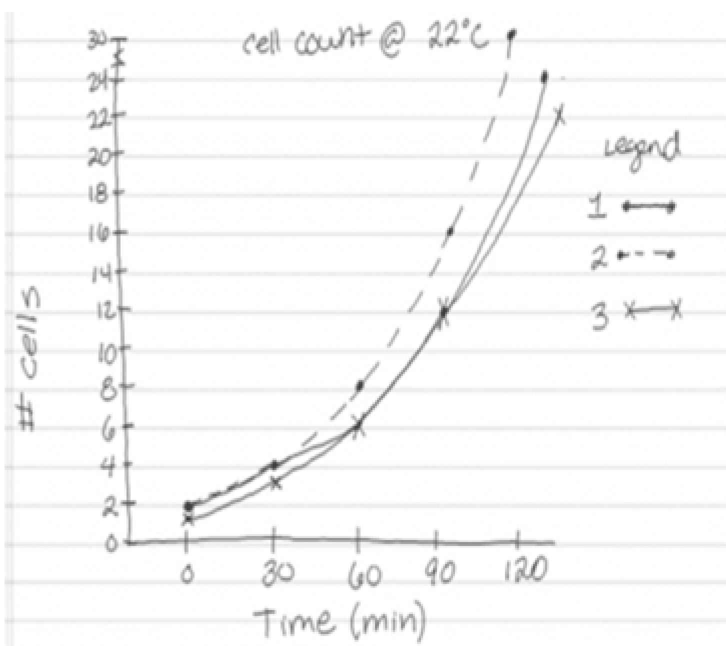
Appendix 2D Continued

Graphs constructed from the bacteria scenario by undergraduates with research experience (UGR)

UGR2_4th year



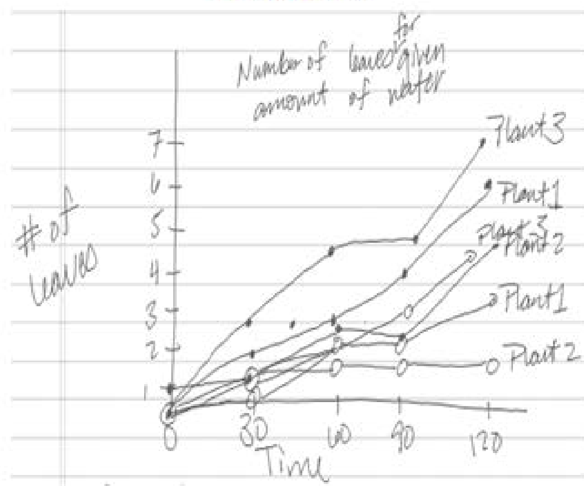
UGR3_4th year



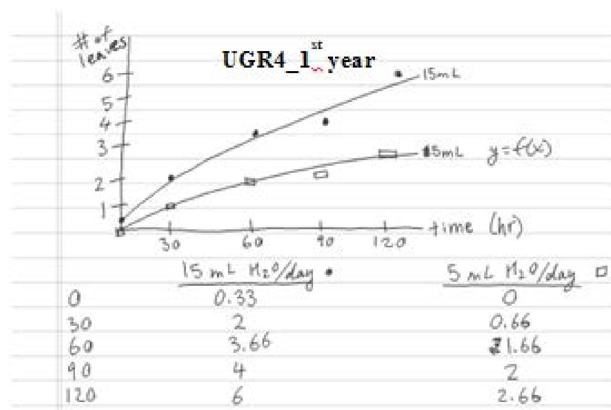
Appendix 2D Continued

Graphs constructed from the plamt scenario by undergraduates with research experience (UGR)

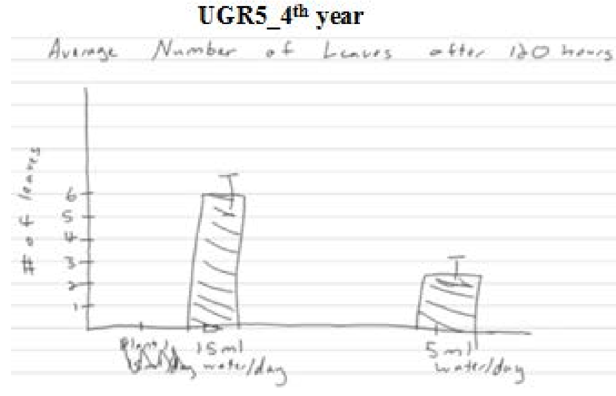
UGR1_4th year



UGR4_1st year



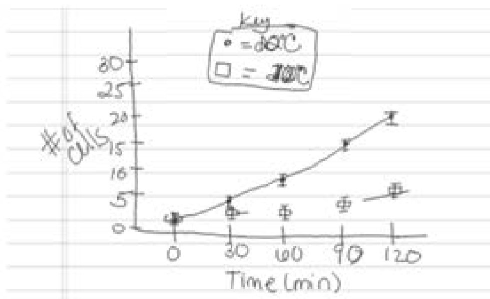
UGR5_4th year



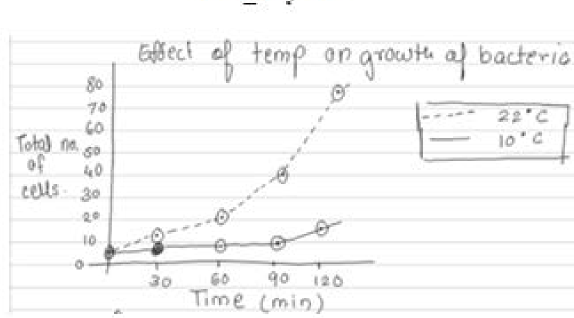
Appendix 2D Continued

Graphs constructed from the bacteria scenario by graduate students (GS)

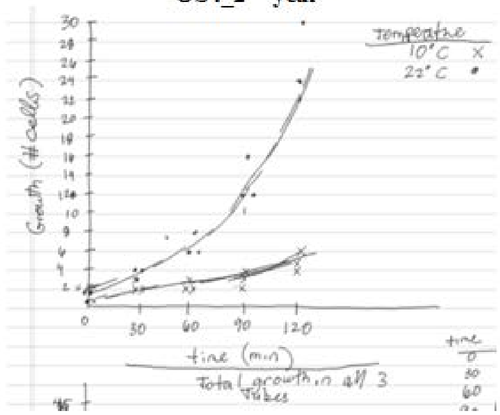
GS1_3rd year



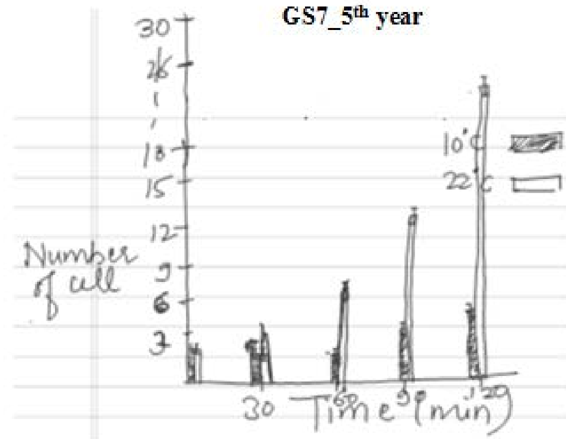
GS6_4th year



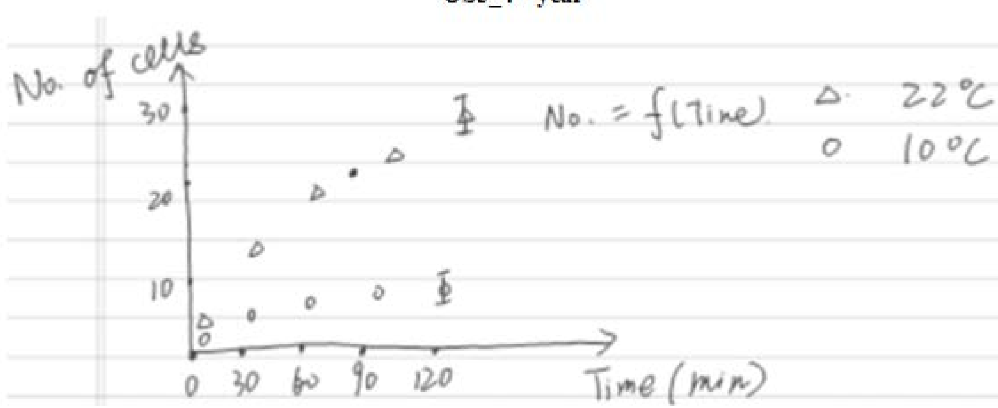
GS4_2nd year



GS7_5th year



GS5_4th year

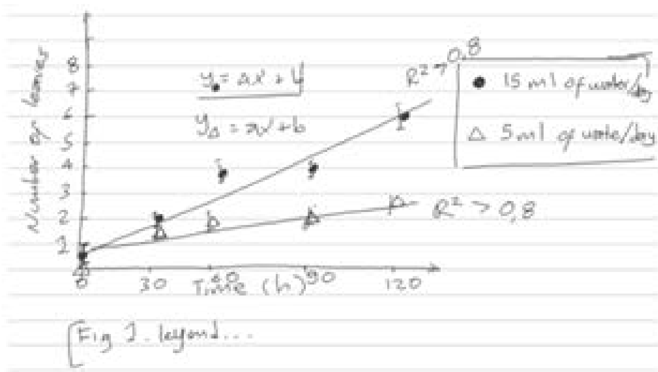


Appendix 2D Continued

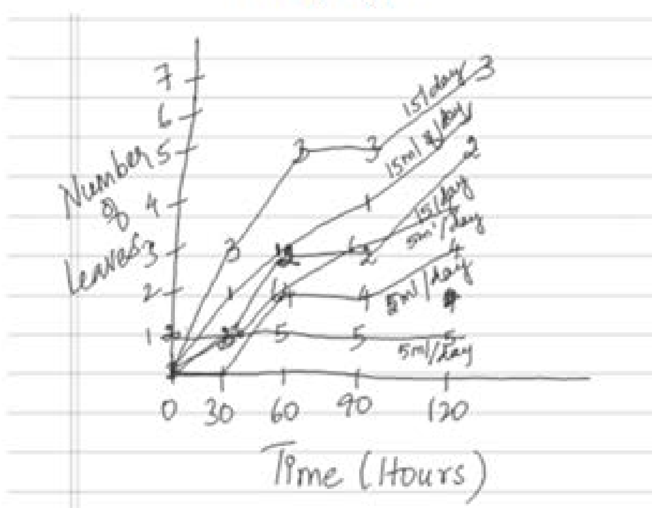
Graphs constructed from the plamt scenario by graduate students (GS)

GS2_4th year

GS3_2nd year



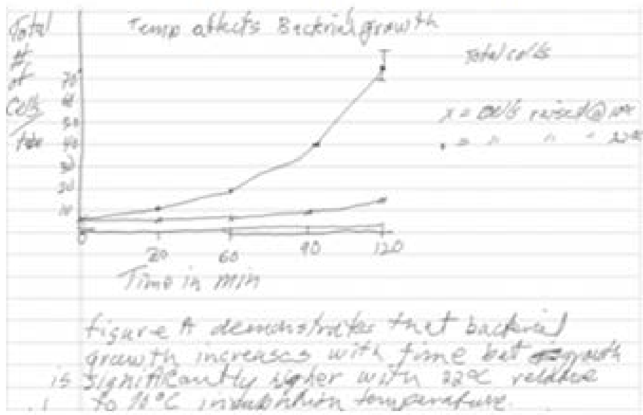
GS8_2nd year



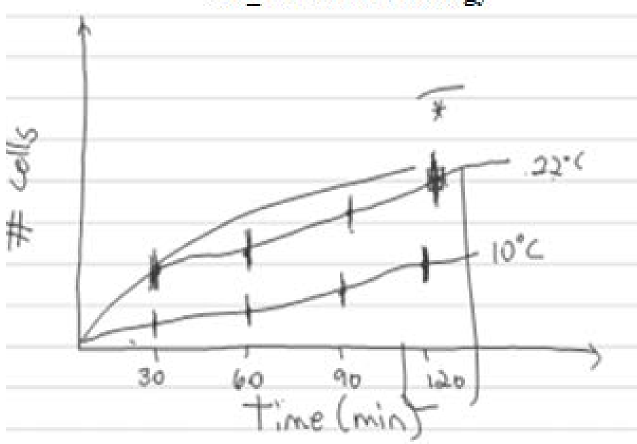
Appendix 2D Continued

Graphs constructed from the bacteria scenario by professors (P)

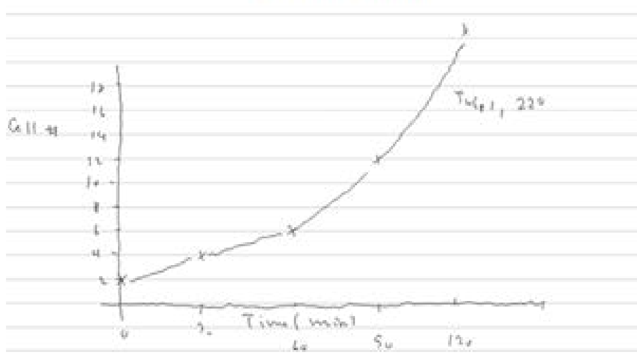
PI_ Behavioral Neuroscience



P2_ Behavioral Ecology



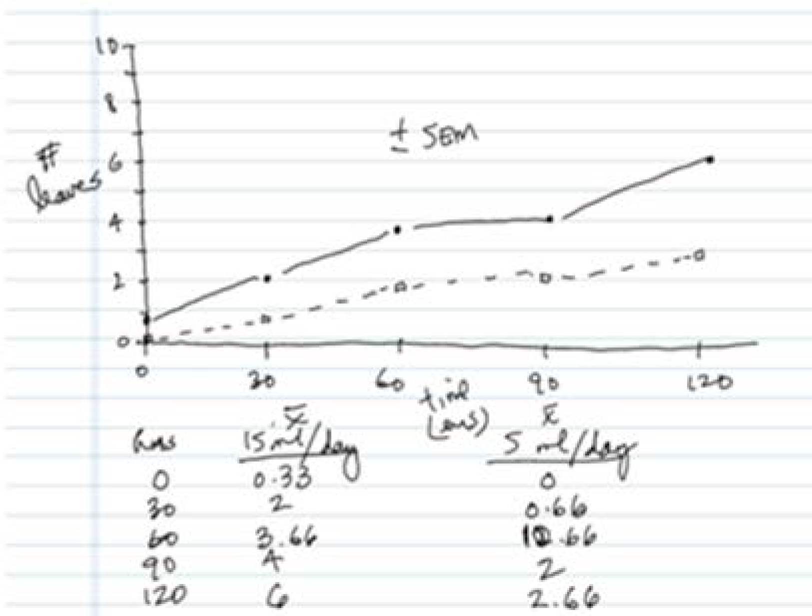
P3_ Microbiology



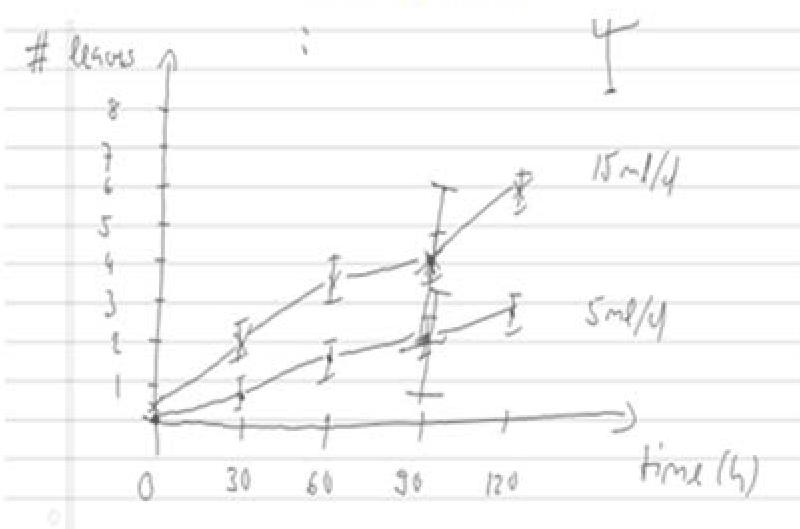
Appendix 2D Continued

Graphs constructed from the plant scenario by professors (P)

P4_Cellular neurobiology



P5_Neurobiology



Appendix 2E: Participants in the Study

Professor's (P) Research Profile and Scenario Given in the Graphing Task

Participant Code	**Scenario**	**Field of Research**	**Teach Graphing in the Classroom?**
P 1	Bacteria	Behavioral neuroscience	All aspects of experimental design and statistical analysis to assist with making relevant choices of numbers of subjects, control groups and inferential statistics appropriate for hypothesis testing.
P 2	Bacteria	Behavioral ecology	Elements of experimental design.
P 3	Bacteria	Microbial genetics & physiology	Enzyme assays with standard deviations and measures of significance.
P 4	Plant	Cellular neurobiology	Discuss experimental data form classic experiments in graphic form for enzyme kinetics and membrane potential chapters. Graphs are used to solve problems.
P 5	Plant	Neurobiology	Discuss graphs and experimental techniques used in neurobiology.

Graduate Students' (GS) Research Profile and Scenario Given in the Graphing Task

Participant Code	**Scenario**	**Doctoral Research Emphasis**	**Year in Graduate School**	**Undergraduate Research Experience?**
GS 1	Bacteria	Microbiology	3rd	UG, screened to identify virulence factors of *Mycobacterium marinum*
GS 2	Plant	Cancer Biology and Immunology	4th	UG, population surveys of freshwater mussels
GS 3	Plant	Plant and Soil Ecology	2nd	UG, characterized magnetotactic bacteria
GS 4	Bacteria	Avian Behavior	2nd	UG, education research, studied task switching
GS 5	Bacteria	Structural Biology	4th	UG, compared phenotypes of wild type and mutant bacteria
GS 6	Bacteria	Virology and Gene Therapy	4th	UG, clinical trials
GS 7	Bacteria	Infectious Diseases	5th	UG, expression, purification and crystallization of a recombinant protein that is involved in the degradation of a specific class of xenobiotics
GS 8	Plant	Retinal Degeneration and Drug Development	2nd	UG, sequenced data from human patients

Undergraduate Students with Research Experience (UGR) and without Research Experience (UGNR) Profile and Scenario Given in the Graphing Task

Participant Code	**Scenario**	**Year in College**	**Major Track**	**Past Research Experience?**
UGNR 1	Bacteria	2nd	General	No
UGNR 2	Plant	2nd	Cell, molecular, development	No
UGNR 3	Plant	1st	Genetics	No
UGNR 4	Bacteria	1st	Genetics	No
UGNR 5	Plant	1st	Biochemistry	No
UGNR 6	Plant	1st	General	No
UGNR 7	Bacteria	3rd	Neurobiology, physiology	No
UGNR 8	Bacteria	1st	General	No
UGNR 9	Bacteria	2nd	General	No
UGNR 10	Plant	3rd	Biochemistry	No
UGR 1	Plant	4th	Cell, molecular, development	Yes, 4 semesters of research, visual images, cell counts, growth rates
UGR 2	Bacteria	4th	Genetics	Yes, results of gels and drosophila crosses
UGR 3	Bacteria	4th	General	Yes, cell count and analysis
UGR 4	Plant	1st	Neurobiology, physiology	Yes, unpaid internship, neuron feedback from cockroaches
UGR 5	Plant	4th	Neurobiology, physiology	Yes, two years of research, population measurements of cellular growth and intensity values

Appendix 2F: IRB Consent form

PURDUE
UNIVERSITY

HUMAN RESEARCH PROTECTION PROGRAM
INSTITUTIONAL REVIEW BOARDS

To:	STEPHANIE GARDNER LILY
From:	JEANNIE DICLEMENTI, Chair Social Science IRB
Date:	09/15/2015
Committee Action:	**Renewal**
IRB Action Date	09/15/2015
IRB Protocol #	1210012775
Study Title	Investigating the reasoning involved in creating graphical representations of data in Biology
Expiration Date	09/14/2016

Following review by the Institutional Review Board (IRB), the above-referenced protocol has been approved. This approval permits you to recruit subjects up to the number indicated on the application form and to conduct the research as it is approved. The IRB-stamped and dated consent, assent, and/or information form(s) approved for this protocol are enclosed. Please make copies from these document(s) both for subjects to sign should they choose to enroll in your study and for subjects to keep for their records. Information forms should not be signed. Researchers should keep all consent/assent forms for a period no less than three (3) years following closure of the protocol.

Revisions/Amendments: If you wish to change any aspect of this study, please submit the requested changes to the IRB using the appropriate form. IRB approval must be obtained before implementing any changes unless the change is to remove an immediate hazard to subjects in which case the IRB should be immediately informed following the change.

Continuing Review: It is the Principal Investigator's responsibility to obtain continuing review and approval for this protocol prior to the expiration date noted above. Please allow sufficient time for continued review and approval. No research activity of any sort may continue beyond the expiration date. Failure to receive approval for continuation before the expiration date will result in the approval's expiration on the expiration date. Data collected following the expiration date is unapproved research and cannot be used for research purposes including reporting or publishing as research data.

Unanticipated Problems/Adverse Events: Researchers must report unanticipated problems and/or adverse events to the IRB. If the problem/adverse event is serious, or is expected but occurs with unexpected severity or frequency, or the problem/even is unanticipated, it must be reported to the IRB within 48 hours of learning of the event and a written report submitted within five (5) business days. All other problems/events should be reported at the time of Continuing Review.

We wish you good luck with your work. Please retain copy of this letter for your records.

Ernest C. Young Hall, 10th Floor - 155 S. Grant St. - West Lafayette, IN 47907-2114 - (765) 494-5942 - Fax: (765) 494-9911

CHAPTER 3. UNDERSTANDING UNDERGRADUATE STUDENTS' REASONING BEHIND GRAPH CHOICE AND CONSTRUCTION IN AN UPPER-DIVISION PHYSIOLOGY LABORATORY CLASSROOM

3.1 Abstract

The Vision and Change and HHMI's Scientific Foundations for Future Physicians documents state that development of graphical competency is an essential skill for students across all disciplines. The purpose of our study is to understand the reasoning implemented by undergraduate students when choosing and creating appropriate graphical representations of physiological data, in an upper-division physiology laboratory course. Four times over the course of the spring 2013 and 2014 semesters, students (n=139) worked in small groups to design experiments relevant to that weeks' topic, collect data, and present findings in a short PowerPoint presentation. After the presentations, students were asked to individually reflect on graph choice, advantages, and disadvantages of their groups' graphical representation. At the end of the spring 2013 and 2014 semesters, student graphs were evaluated qualitatively based on four categories: graph mechanics, data form, graph choice, and aesthetics and visuo-spatial considerations, and the reflection responses were coded using thematic analysis. Findings reveal that the most common types of graphs constructed were: bar, scatter, or line graphs, and dot plots and box and whisker plots were the least common. Student

reflections for graph choice fell into five main themes: technology, graph interpretation, communication, reflection, and experimental concepts. There was no explicit mention of the question or hypothesis in the students' reflections on graph choice. Students often correctly identified the advantages, associating variable type with graph type. However, when articulating the disadvantages, students often stated the advantages or self-critiqued their graphs. Findings from this study influenced the development of several graphing instructional materials to improve student reasoning with graph choice and construction.

3.2 Overview

There have been numerous calls to improve the undergraduate biology curriculum pertaining to incorporating mathematics and inquiry into the classrooms in the STEM disciplines (Labov et al. 2010; Brewer & Smith, 2010; Gormally et al., 2009). In 2011, the *Vision and Change in Undergraduate Biology Education* document (AAAS, 2011) listed "ability to use quantitative reasoning" as one of the core competencies that all undergraduate students should achieve to use not only in the classroom, but also to evaluate outside information in their daily lives (AAAS, 2011). Skills associated with graph construction and interpretation are a subset of quantitative reasoning and are recognized by faculty teaching general biology courses as skills necessary for science literacy (Gormally, Brickman, & Lutz, 2012).

Although graphing instructional books (Bertin, 1983; Tufte, 1983; Kosslyn, 1994; Few, 2004) and online tools (Create A Graph Tutorial-NCES Kids' Zone, 2016; Interactive Statistics Map, 2016) exist to aid students (Webber et al., 2014) with graph construction, they provide very general feedback to the graph constructor, perhaps because these materials are not situated in a specific disciplinary context. This may contribute to the lack of validated tools for graph construction in science is because the reasoning that occurs when working with scientific experimental data in a classroom setting has not been studied in depth. This is the precise focus of the work described in this chapter which is to utilize findings from student reflections to inform the development of graphing materials to be used during instruction in a classroom laboratory setting (chapter 4). Findings presented in this chapter will complement findings reported in chapter 2 of the dissertation, where think-aloud interviews were conducted with expert professors, graduate students, undergraduate students with and without research experience (novices) to understand how they reason with graph choice and construction. Since the study in chapter 2 was in a clinical setting and participants were given generic data, it did not emulate conditions in which we want to use our graphing materials. Thus it was necessary to complement the findings from the expert-novice study in chapter 2, with findings from student reflections on data that they collected from their experiments, in a classroom setting.

3.3 Background

Successful construction and reasoning with external representations, such as graphs, requires students to be knowledgeable and reflective in four areas: invention, critique, functioning, and learning or reflection. These four components comprise meta-representational competence, or MRC (diSessa & Sherin, 2000; diSessa, 2004) which describes the knowledge and skills of a competent representation maker. The first component, invention, explains how students are able to conceive novel graphical representations from data (diSessa & Sherin, 2000). The second component, critique, exposes students' critical knowledge for assessing strengths and weaknesses of various types of graphs (diSessa & Sherin, 2000). The third component, functioning, explains students' underlying reasoning for connecting the function and usage of various graphs on the type of data present (diSessa & Sherin, 2000). The final component, learning or reflection, is a strategy that fosters students' awareness of their own understanding of graphs and gaps in knowledge (diSessa & Sherin, 2000). Self-reflection and critical thinking are some of the best ways for students to evaluate and think about their learning (Tynjala, 1999). Although oral reflections via interviews are a rich source of data to convey to the researcher what the student is thinking, they are oftentimes not practical in a naturalistic setting. Written reflections are noteworthy when they probe students to think deeply. In order to increase students' confidence and refine their critical thinking skills, and the learning or reflection component of the MRC, reflections should be performed and encouraged numerous times throughout a class.

A study involving student reflections was done in the context of an introductory physics course, where students were asked to submit weekly report, reflecting on the physics they learned (May & Etkina, 2002). The study reported that students who had the best grasp of the physics concepts also had more articulate reflections than students whose reflections were of lower quality (May & Etkina, 2002). This study shows that reflections do provide insight into the student's mind and their thought processes. In the context of undergraduate biology, there have been two studies that provide insight into student difficulties with data and graphing (Bray-Speth et al., 2010; McFarland, 2010) McFarland (2010) designed a laboratory class intervention devoted to graph choice and construction that included instructions on graphs, their usage, and student engagement in the reflection on appropriateness of graphs. Over the semester, students were asked to reflect on their graph choice and quality, engaging the learning or reflection component of MRC. Although students were not explicitly assessed on their graphs or reasoning, students' graphing quality improved over time and students responded positively to a course evaluation question about their learning about graphs in the class. Unlike the McFarland (2010) study, the Bray- Speth et al. (2010) study occurred in the context of an introductory undergraduate biology lecture class. Students showed significant improvements in graph mechanics (title, axis labels, units, scale, and key), but there was an inconsistency in graph choice. Although these studies report interesting and useful findings in the undergraduate biology context, it is challenging to deduce best practices and instructional tools to promote the development of graphing proficiency, without having an understanding of student thinking with graph choice and construction. Our study aims to fill this gap in our insight regarding student understanding with graphing by providing vital information from student reflections from two semesters (n=139), which will inform and improve instructional approaches as they are relevant to graphing in the biological context.

3.4 Theoretical Framework

The research is guided by the constructivist theoretical framework, which states that individuals construct new ideas from their past knowledge and experiences (Duffy & Jonassen, 1992). In constructivism, individuals do not discover new knowledge but they actively construct it with their existing concepts and models and then modify their understanding with new experiences (Bodner, 2006; Bodner & Orgill, 2007). The constructivist theoretical framework is an ideal as an overarching framework for this study because we aimed to understand the foundation from which our students are constructing new knowledge (Fosnot & Perry, 2005; Cooperstein & Weidinger, 2004). Going forward, we plan to use the knowledge exhibited by students in this study, build on it by informing the development of graphing materials, so that students in the future can reap the benefits.

3.5 Research Questions

The overarching research objective of this study was to utilize reflections from physiology students enrolled in a Principles of Physiology course as a means to understand their reasoning behind graph choice, characterize the attributes of their graphs, and to help future students. Findings ultimately informed the development of graphing instructional and learning materials. To accomplish this objective, we sought to answer three questions:

1. How do undergraduate students enrolled in an upper-level physiology laboratory reason with graph choice for their experimental data?
2. How do undergraduate students enrolled in an upper-level physiology course reason with advantages, and disadvantages to display the experimental data they collected from experiments they designed?
3. What types of graphs are created by undergraduate students enrolled in an upper-level physiology course and what are the attributes of their graphs?

3.6 Methods

3.6.1 Participants

This study was conducted in the laboratory portion of an upper-level physiology course (Principles of Physiology) at a large midwestern, research university during the spring 2013 and 2014 semesters. Students enrolled had completed calculus for life sciences or cell structure and function, as prerequisites for the physiology course. The laboratory course aligned with the course material taught in the lecture portion of the course and was designed by the instructor, hereafter referred to as SMG. All laboratory experiments were inquiry-based, open-ended, and allowed students to work together in teams of four to design experiments, collect data, and present data in a 10 minute PowerPoint presentation. This process occurred five times over the course of each of the 2013 and 2014 semesters, but we report data from four labs (Figure 3.1) due to time constraints that did not allow students to fill out the individual reflection.

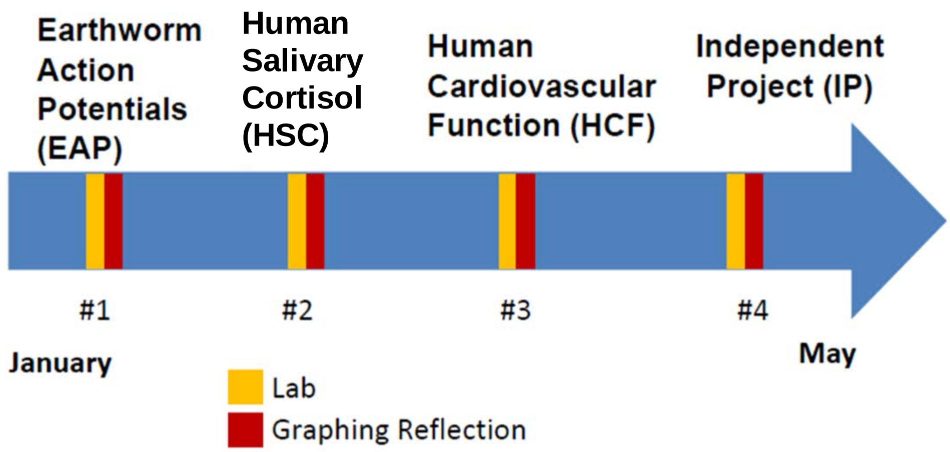


Figure 3.1: Timeline for Spring 2013 and 2014 Semesters: The timeline above illustrates the four labs from which we collected reflection data focusing on graph choice, advantages, and disadvantages.

Under an exempt IRB protocol (#1208012562, see Appendix 3H), we report on graph reflections and graphs constructed from data collected in four labs (Figure 3.1) by 139 students over the course of the spring 2013 and 2014 semesters. At the time of study, the majority of the students enrolled in the 2013 and 2014 semesters were either pursuing bachelors in biology in neurobiology and physiology or general biology (See Appendix 3A for detailed demographic information). Although the majority of the students enrolled in the course were rising juniors or seniors, we assumed all physiology students were at the journeyman level when they signed up for this course. This term was coined by Bing and Redish (2012) and it refers to students who have the fundamental knowledge and skills that does not make them novices, but their lack of practice with problem solving also does not categorize them with the experts. Therefore, these students are assumed to be midway between novice and expert, thus the term journeyman.

3.6.2 Laboratory Context

Each class laboratory session started out with a short PowerPoint reinforcing the concepts learned in lecture prior to lab and providing a brief overview of the lab objectives, equipment used, as well as the expectations from students. At the beginning of the semester students were assigned to permanent small groups with the help of a tool developed by Ohland and colleagues (2006), called CATME (Comprehensive Assessment of Team Member Effectiveness). On experiments days, students were expected to work together to formulate research questions, hypotheses, design experiments, collect data, and present findings in a short 10 minute PowerPoint presentation to the class. There was no formal instructional unit on graphing, but rather SMG and/or Aakanksha Angra (hereafter referred to as AA) were present in all three of the lab sections to provide just-in-time teaching, assisting students with their experiments, data analysis and graphs as needed. In addition, for the entire semester, students had access to a graphing primer designed by SMG that they could access at any time through Blackboard Learn (See Appendix 3B). Students were expected to contribute equally to the group work. SMG and AA kept field notes on the groups they helped and the type of feedback they gave to the students and met weekly to discuss their notes and observations. After each experimental lab, students gave presentations to the class and immediately following, they either independently handwrote (Spring 2013) or typed (Spring 2014) their reflection on two questions:

1. *Why was the graph(s) you used chosen over other graphical options?*
2. *What are the advantages and disadvantages of this representation(s)?*

To ensure high quality of student answers to these questions, point values (5 points) were assigned for completion. Students were not given feedback on their graph reflections, but were provided feedback on their experimental design, graphs and oral presentations. A coarse-grained rubric (see Appendix 3C) was used to provide students feedback on their graphs and figures, presentation, and oral communication. With regards to their graphs, the main areas where students received feedback were: appropriate display, which refers to graph choice; labeled axes and title that should be descriptive for the data displayed; descriptive and inferential statistics, and visual accessibility, referring to the aesthetics of the graph. For detailed definitions, please refer to Appendix 3C. All written student responses, PowerPoint presentations, and graphs were de-identified to maintain confidentiality.

3.7 Data Analysis and Coding

3.7.1 Student Reasoning with Graph Choice

In order to answer the first research question, the following steps were taken to prepare data for coding student responses were: converted to a digital format, assigned a number to maintain confidentiality, and organized by the type of graph, the reflection question, and the lab. Inductive and deductive analysis was used to organize and code the student responses (Patton, 2005; Strauss & Corbin, 1998; Saldana 2013). Codes were identified per the definition provided by Saldana (2013) as "*a word or short phrase that symbolically assigns a summative, salient, essence-capturing, and/or evocative attribute for a portion of language-based or visual data*". This type of analysis was appropriate for our data because we used words, verbatim from the students' reflections in order to obtain an understanding of their reasoning with graph choice. Our coding scheme went through numerous iterations over the span of Summer 2013-Fall 2014, until it was finalized and validated. In the first cycle of coding, both SMG and AA independently memoed the transcripts in order to get a basic understanding of what students were saying. Although some of our codes were inspired by the students' reflections themselves, we also took a deductive approach and consulted numerous sources in the graphing literature to inform our initial coding scheme (e.g. Kosslyn, 1994; Friel & Bright, 1996; Schriger & Cooper, 2001; Cooper et al. 2002; Patterson & Leonard, 2005). In the second cycle of coding, we looked for patterns and grouped students whose reflections were similar to one another and those who reflected on the same type of graph. In our third iteration, selective coding was used to organize the codes into a story that described the complex network of themes that emerged (Creswell, 2013). Several more iterations occurred and data from student responses were recoded, recategorized, and refined. Our final coding scheme consisted of five themes with primary and secondary codes embedded within each theme (Appendix 3D). Both SMG and AA independently coded at each step of the coding process, met regularly to compare and discuss the coding. Instances during which there was disagreement, discussions were held, until a consensus was reached on the final themes. The final themes were validated by multiple colleagues (faculty and graduate students in discipline-based research) attending the Purdue International Biology Education Research Group meetings. The five final themes that comprised our coding scheme are described below.

Technology: This is when a student mentioned using computer software to construct their graph. We chose to classify *technology* as its own theme since students in the laboratory were trained on and asked to construct their graphs using Microsoft Excel or Origin Pro, we saw students express their concerns in these graph reflections. For example, one student stated th*at "This was the only type of graph I knew how to make in Excel"*. Furthermore, past studies have documented student engagement with graph construction using graphing software (Leonard & Patterson, 2004; Patterson & Leonard, 2005; Patterson & Leonard, 2007), which also justified our classification of this theme. See Appendix 3D for examples.

Graph interpretation: The definition for this theme was taken from Padilla, McKenzie, and Shaw's work (1986), "*Proper graph interpretation includes: being able to define the relationship between variables; ability to reasonably interpolate and extrapolate; compare and relate the results of a given graph to natural phenomenon in question"*. The first primary code that emerged from student reflections was *comparison*, with the following categories being secondary codes: *between subjects, within subjects, between data points*, *between measured variables*, and those who expressed their *personal opinion*, for example noting that "*it was easy to compare XYZ*" (Elliot et al., 2006; Bruno & Espinel, 2009; Trolier & Hamilton, 1986, Tairab & Al Naqbi, 2004; Shah, Meyer, and Hegarty, 1999). The second primary code that emerged from student reflections was *difference* or *change*, with the same secondary codes as listed above for the primary code, *comparison* (Elliot et al., 2006; Bruno & Espinel, 2009; Trolier & Hamilton, 1986, Tairab & Al Naqbi, 2004; Shah, Meyer, and Hegarty, 1999). The third primary code was *relationship*, with secondary codes *measured variables*, and *personal opinion* (Millar, 2001; Ates & Stevens, 2003; Elliot et al., 2006; Bruno & Espinel, 2009; Trolier & Hamilton, 1986, Tairab & Al Naqbi, 2004; Padilla, McKenzie, and Shaw, 1986; Shah, Meyer, and Hegarty, 1999). The next primary code was *statistical concepts*, with secondary codes of *descriptive and inferential*, which we saw emerge both in our data and literature (Millar, 2001; Trolier & Hamilton, 1986; Hattikudur et al., 2012; Picone et al., 2007). The final primary code was *take-home message*, which consists of formulating the main message the graph is conveying (Shah, Meyer, and Hegarty, 1999). See Appendix 3D for examples.

Communication: We defined this theme as commenting on the appearance, aesthetics, graph quality, and graph functions. There were three primary codes and eight secondary codes that fell under this theme. The first primary code is, *figural effects* (diSessa, 2004), which states that students pay attention to aesthetics and Gestalt rules such as proximity of lines and offering their personal opinion on the quality of the graph, instead of focusing on the function of the representation. The second primary code is, *function* which is part of MRC and it describes the purpose and usage of the graph (Schriger & Cooper, 2001; Duke et al., 2015; Angra & Gardner, 2016). Within the category of function, there were two sub-codes: *data* (Konold et al., 2015) and *statistical concepts* (College Board, 2012). The sub-codes *data* and *statistics* are taken from the work done by Konold et al., (2015), who state that data can be viewed through multiple lenses: pointers, case values, classifiers, and aggregates. Each of these data lenses can be used to organize data differently and to answer different questions (Konold et al., 2015). Although there is a hierarchy in these lenses (see Figure 3 in Konold et al., 2015) the aggregate data lens is the most complex and requires knowledge of statistics (descriptive or inferential) (Konold et al., 2015). The last primary code is, *graph components* (Åberg-Bengtsson, L., & Ottosson, 2006;Schriger & Cooper, 2001; Cooper et al 2002; Kosslyn, 1994; DiFazio, 1990; Elliot et al., 2006; Leonard & Patterson, 2004; Patterson and Leonard 2007; Ainley 2000; Brasell & Rowe 1993) which consists of three secondary codes, *variables, describing the layout of the graph,* and *linking the data plotted to the graph.* See Appendix 3D for examples.

Reflection: This theme is also part of MRC (diSessa, 2004) and consists of two primary codes: *suggestions for improvements* and *other graphs*. The first primary code, suggestions for improvements refers to the student reflecting on the effectiveness with which their graph is displaying data. The second primary code, *other graphs* is when students used other types of graphs to justify their reflection on their graph choice. This code is further broken down into secondary codes: *generic, other graph types, aesthetics, data,* and *variables.* The secondary code *generic* is when students did not specify a graph option, but made an overall statement of *"Other graphical options did not allow us to* *present the data in the most effective manner”.* The secondary code *other graph types* are when students explicitly stated another type of graph than what they were reflecting on. The next secondary code *aesthetics* is when students linked another graph type to its *aesthetics*. The next secondary code, *data* is when students described data as qualitative, quantitative, categorical or continuous, but associated with another graph that they were not reflecting on. The last secondary code is *variables* and this is similar to the previous code but instead of linking data to other graphs, students explicitly stated the word *variable*. See Appendix 3D for examples.

Experimental Concepts: We included this theme in our coding scheme because students were reflecting on their graphs from experiments they designed, so we expected that they would mention terms associated with designing and conducting experiments. There are six primary codes: *research question/purpose, hypothesis, data collection/manipulation*, *limitations*, *variables,* and *personal opinion*. The primary codes *research question* and *hypothesis* were included in the coding scheme to see if students aligned their graph choice to the research question and hypothesis. The next primary code, *data collection and manipulation* was to see if students talked about how they collected data and whether or not if they transformed their data (Konold et al., 2015). The next primary code, *limitations* refers to students articulating limitations of their experiment. The secondary codes embedded within limitations are *time* and *samples.* When talking about *limitations*, students reflected on *time* constraints and a small *sample* size for their graph choice. The next primary code, *variables* is when students were explicitly using the word “variable”, and connecting it to their experiment. Lastly, we included a primary code for *personal opinion*, if students chose to state how they felt about their experiment and graph. See Appendix 3D for examples.

Each theme in the coding scheme contains elements that are repeated because we wanted to stay authentic to data from student written reflections so we could accurately report how students were reflecting on their graphs.

A final table summarizing the five themes (technology, graph interpretation, communication, reflection, and experimental concepts) with citations and examples can be found in Appendix 3D. An example of how a student's reflection was coded is displayed in Appendix 3E.

Since the purpose of this study was to understand how students reason with graph choice and construction to inform graphing instructional and learning materials, we do not report statistics for these data but we do quantify them (Figure 3.2) in order to quickly generalize patterns in order to acknowledge the presence or the absence of themes across different types of graphs and years the data was collected. Although we agree that context may influence student performance (Kanari & Miller, 2004) or that students may improve their reasoning with multiple rounds of practice (Roth & McGinn, 1997), we are grouping data and reporting findings on graph type. More detailed analysis of some of these data is summarized in chapter 5 of this dissertation.

3.7.2 Student Reflections on Advantages and Disadvantages of their Graphs

In order to answer the second research question, looking at how students reflect on the advantages and disadvantages of their graphs, we created word clouds to analyze student responses because they are rich visuals that allow quick analysis of qualitative data (McNaught & Lam, 2010). Since data from this chapter are informing the development of graphing instructional and learning materials (chapter 4), we found this method useful as an informative tool. The most common words used are represented by large font and least common words are represented in small font. All word clouds were made online using a free website (WordClouds.com). Before generating word clouds, student responses for the second reflection question were divided into advantages and disadvantages. Data were combined for all labs, but years were kept separate from one another because we were not evaluating student responses based on lab topic context, but we did want to see if similar words were present for student reflections in both Spring semesters. Words such as "bar, line, scatter, box, dot, and graph" were deleted because in a preliminary word cloud analysis, these were the largest words and we were interested in other things students were saying. We also combined words such as "comparison, compare, comparing, compared, and compares" into the same tense. This was followed for other words as well. Examples of word clouds for bar graph are displayed in Figure 3.4 and all other word clouds can be found in Appendix 3F, Figures 3A-F.

We also performed a cursory analysis of the correct and incorrect reasoning uttered by students in their reflections to get a sense of what was being said for each graph type.

3.7.3 Characteristics of Student Graphs

To address the third research question, graphs constructed by students were removed from their PowerPoint presentations, along with the research question and hypothesis and pasted into a Microsoft Word file. We qualitatively described the graphs based on four broad categories: graph mechanics, data form, graph choice, and aesthetics.

The detailed list of the evaluation categories are listed in Chapter 2 of this dissertation, Appendix 3C.

3.8 Results

We used qualitative methods to analyze student written reflections and graphs to answer our three research questions:

1. How do undergraduate students enrolled in an upper-level physiology course reason with graph choice for their experimental data?
2. How do undergraduate students enrolled in an upper-level physiology course reason with advantages, and disadvantages to display the experimental data they collected from experiments they designed?
3. What types of graphs are created by undergraduate students enrolled in an upper-level physiology course and what are the attributes of their graphs?

3.8.1 Reasoning with Graph Choice

To address our first research question, looking at students' reasoning with graph choice, we report reflection data (n=556 reflections) from 139 undergraduate students that were coded via inductive and deductive analytial methods. The five themes that emerged from students' reflections were: *technology, graph interpretation, communication, reflection,* and *experimetnal concepts* (Appendix 3D). Figure 3.2 displays the themes that emerged for graph choice in the form of stacked bar graph with year and graph type on the x-axis and percentage of occurrences of themes on the y-axis. The percentage of occurances were calculated by first counting the number of instances of a code that appeared for each theme, within the year, dividing it by the total occurances of themes for that year, and multipying it by 100.

Student responses fell mainly within the themes of *graph interpretation* and *communication*. The theme technology only appeared once, when students were reflecting on their graph choice of dot plots, which were one of the uncommon graph types created (see Figure 3.5). Although we expected students to talk about *experimental concepts*, since their graph stemmed directly from the data they collected, we didn't see see this as a common theme and was completely absent in students' reflection for the box and whisker plots in 2014. However this could have been due to the small sample size (n=23) of box plots created in total. Like the theme *experimental concepts*, the theme *reflection* was equally uncommon in student reflections.

Since the themes *graph interpretation* and *communication* were the most common themes, we looked into each one at a finer scale. Appendix 3G shows the primary codes for the themes *graph interpretation* and *communication.*

Looking broadly at the primary codes for *graph interpretation,* we see a greater presence of the primary code comparison, differences/change. The primary code, *take-home message* was the least common across all different types of graphs. The primary code, *relationship* was the most prevalent in scatter plots, while the primary code *statistical concepts* was popular in both scatter plots and box and whisker plots.

Next, looking at the stacked bar graph (Appendix 3G) of the primary codes for *communication*, we see that the most common primary code was *function*, followed by *figural effects*, and then *graph components.*

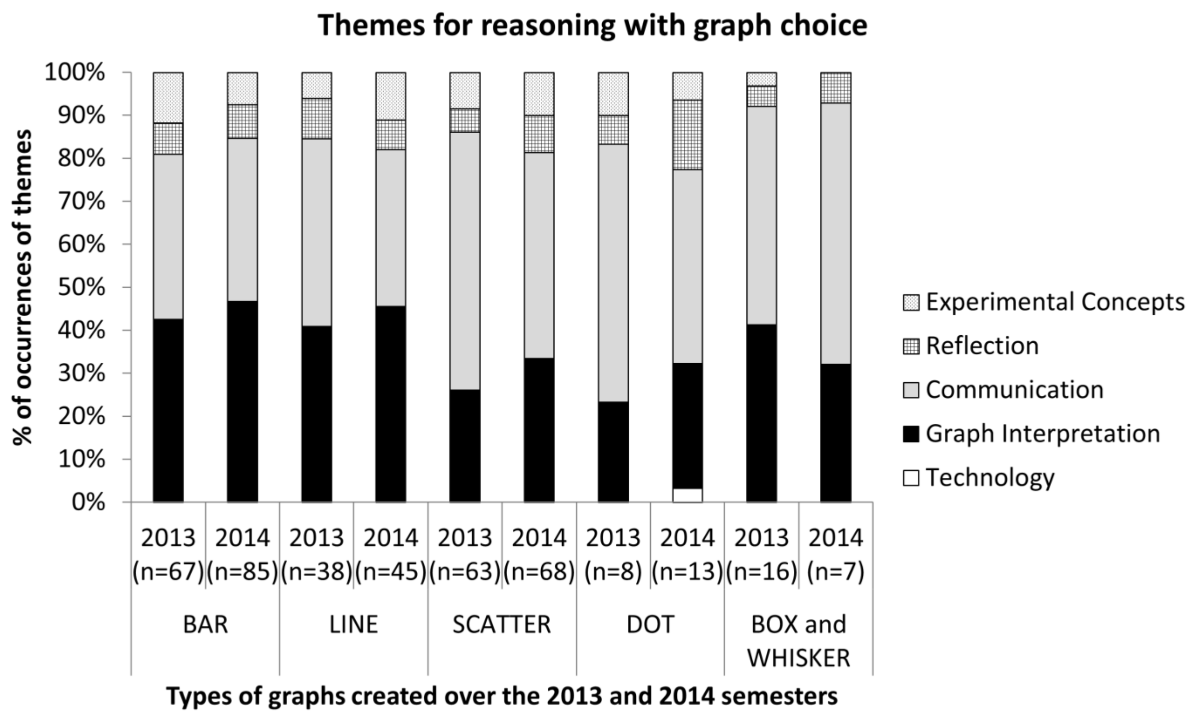


Figure 3.2: Themes for reasoning with graph choice: Types of graphs created over the 2013 and 2014 Spring semesters and the percentage of occurrences of themes for each type of graph within the semester of interest.

3.8.2 Reasoning with Advantages and Disadvantages

To address our second research question, looking at students' reasoning with advantages and disadvantages, we used two methods of data analysis. The first method was cursory, in order to understand how students were reflecting on the advantages and disadvantages and the second method consisted of synthesizing data through word clouds. Here we provide some examples of student answers to the advantages and disadvantages of their graph type (Figure 3.4). We noticed that most students did not show difficulty when reflecting on the advantages, linking bar graphs with categorical variables and line graphs with the continuous variable, time. However some students also showed uncertainty as to what constituted a "disadvantage" because they reflected on the purpose and actually listed an advantage instead of disadvantage. Another interesting observation (part of our brief analysis) was that if student groups created multiple types of graphs, ranging in style and complexity of data plotted, not all group members reflected on all graphs, but chose the ones they felt comfortable reflecting on.

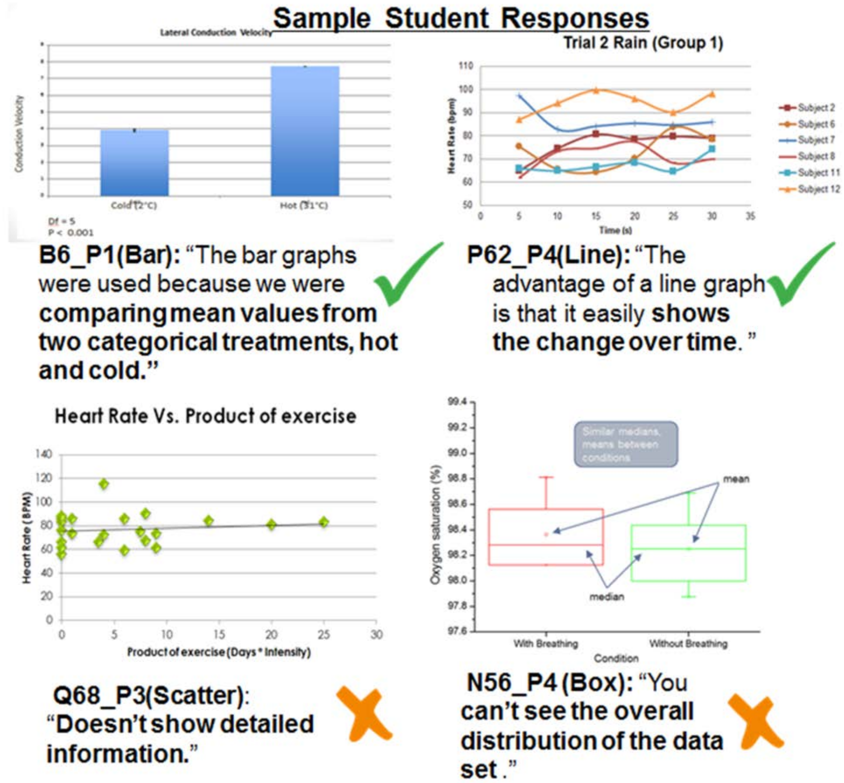


Figure 3.3: Examples of student graphs with reasoning for advantages and disadvantages: Figure shows students' correct responses to the advantages of their graph (top) and incorrect disadvantages (below). This figure is representative of the types of responses students gave, ie. students more often and correctly provided the advantages of their graphs but more often provided incorrect disadvantages of their graphs. Although we did not quantify the degree of "correctness", it is a future step in data analysis.

The second method that we used to understand student reflections on advantages and disadvantages was by using word clouds. Across all the graphs reflections, we see a high frequency of words like, “data” and “show”, which tells us that students have an understanding of the general purpose of graphs, which is that they serve as communication tools because they show data. We also noticed words such as “easy” and “clear” appeared across all graph types for the advantages and words like “difficult” and “confusing” appeared for the disadvantages, which illustrate graph quality and personal opinion of the student. Students also frequently incorporated words from their experimental design to explain the advantages and disadvantages of their graph choice. Students also used words specific to each graph type that illustrated the function of that graph. For example, in the word clouds shown in Figure 3.4, we see students use the words like “categories” and “comparison” when talking about the advantages and the disadvantages of bar graphs. We also see students describe the type of data plotted in their bar graphs, indicated by word like “averages”. In Appendix 3F, we see that students commonly used the word “trends” when talking about the advantages of line graphs. Some common words that students associated with the advantages of a scatter plot (Appendix 3F) were “correlation”, “trend”, “relationship”, and “regression”. With box and whisker plots (Appendix 3F), students used words associated with descriptive statistics to point out that box and whisker plots show the “average”, “median”, “maximum”, and “range” of the data they collected in lab. For dot plots, we noticed students used words like “category”, “individual, and “distribution” when describing the advantages.

Word clouds for the disadvantages of the graphs closely resembled those of the advantages. Missing from all word clouds were students connecting their graph to their research question and hypothesis. Word cloud analysis is a quick way to analyze qualitative data and it will inform our future instruction with graphing and allows us to understand what types of information we want students to know and incorporate into their responses.

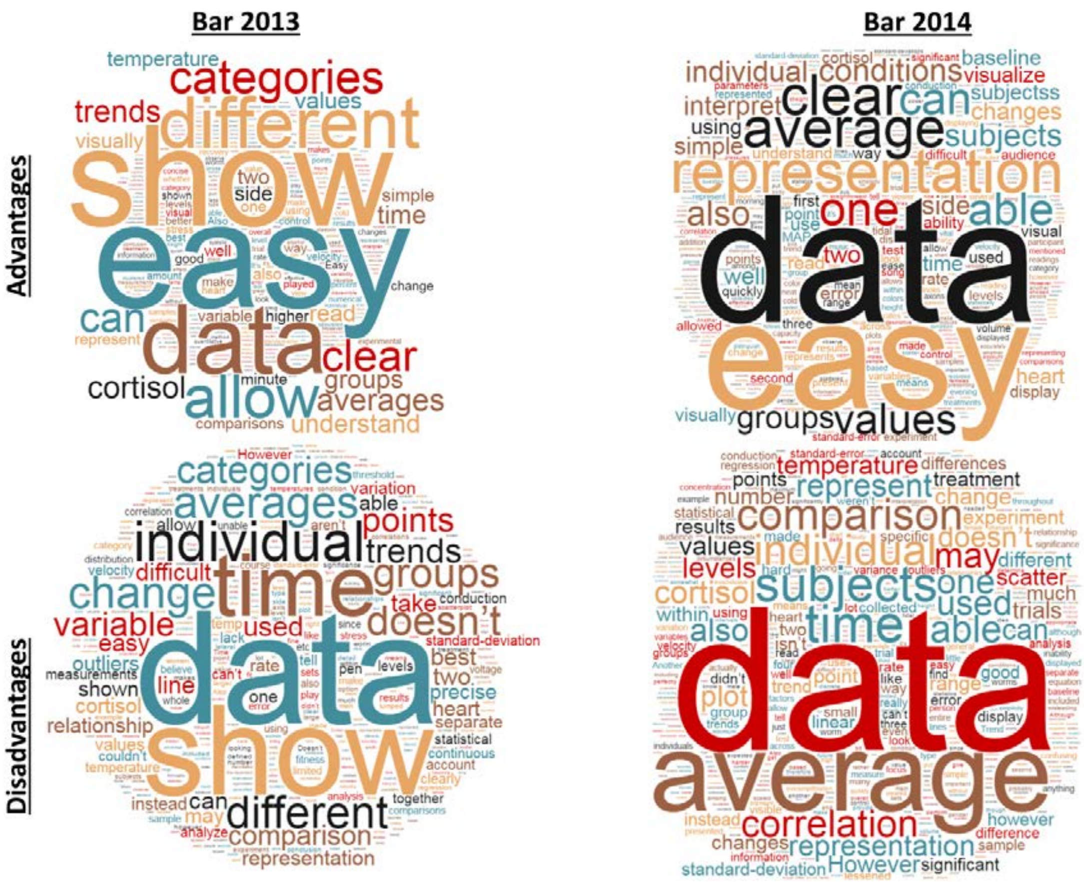


Figure 3.4: Word frequency of advantages and disadvantages of bar graphs: Figure shows a side by side comparison of word frequencies of the advantages and disadvantages of student reflection on bar graphs.

3.8.3 Types of Graphs Constructed

To address our third research question aimed at characterizing the attributes of graphs constructed by groups of students, we categorized the graphs qualitatively. Using the same criteria as used in chapter 2 of this dissertation, we examined graphs as stand-alone artifacts, by extracting them from students' PowerPoints. First we separated the graphs by graph type for each lab over the course of the semester to see if there were differences in the types of graphs constructed over the course of the Spring semesters (January-May).

Findings from both Spring 2013 and 2014 semesters are illustrated in a series of pie charts in Figure 6. It is interesting to note that in both the Spring 2013 and 2014 semesters, there were fewer types of graphs constructed in the beginning of the semester and a larger variety of graphs constructed by the end of the semester. There was a higher percentage of bar graphs in the beginning of the 2013 semester and it gradually decreased (60% to 19%), whereas in Spring 2014, the percentage of bar graphs stayed close to one-third of the total graphs constructed, where in the BP/EKG lab, it was almost 50%. The percentage of scatter plots constructed in Spring 2013 stayed close to one-third of the total graphs constructed, whereas in Spring 2014, we see that in the first lab, scatter plots made up the majority of graphs constructed, with 56%, and decreased to 28% in the last lab of the semester.

The patterns for line graphs constructed were very similar across semesters and labs. In the second lab of the semester (HSC), we see the presence of dot plots and box and whisker plots. Five percent of graphs constructed in 2013 for the BP/EKG lab were box plots, whereas 9% of graphs constructed were dot plots for the same lab, but in Spring 2014.

For the last lab in Spring 2013, we see the appearance of pie charts (3%), along with dot plots (6%) and box and whisker plots (15%). In Spring 2014, we only see box plots 6% and the other 94% percent are bar, scatter, or line graphs.

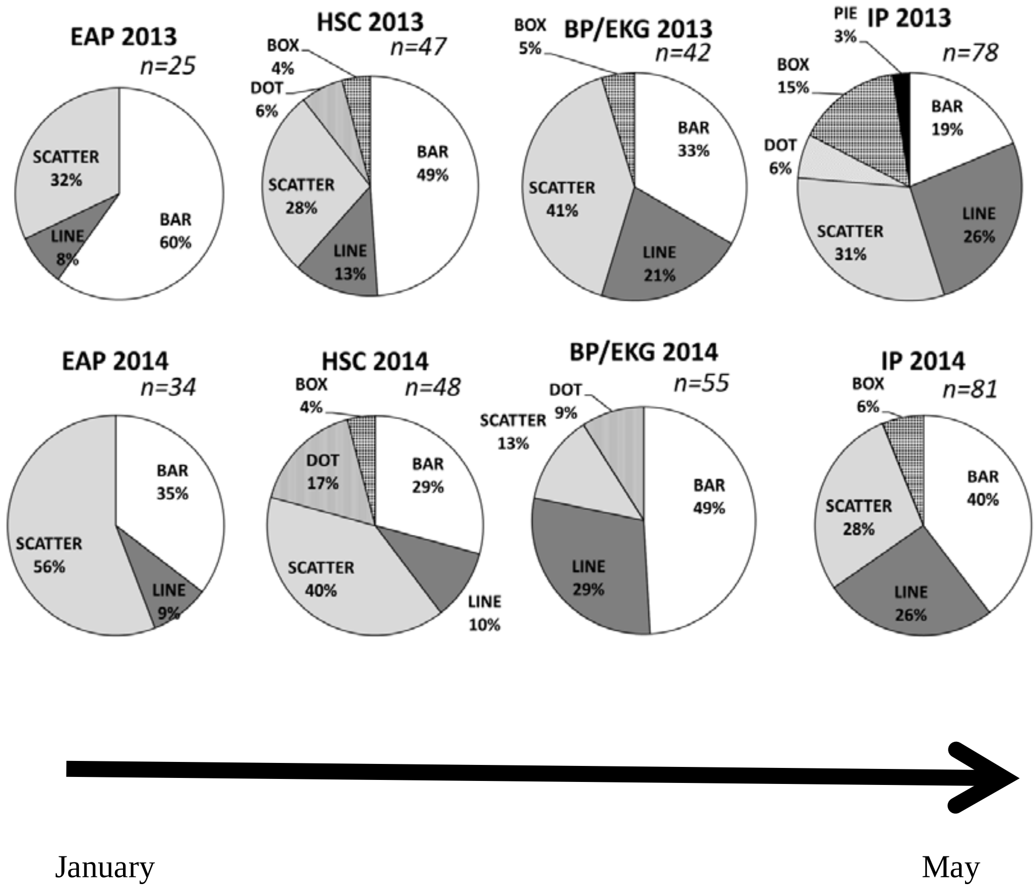


Figure 3.5: Types of graphs constructed in Springs 2013 and 2014: The series of pie charts display the types of graphs constructed for each lab and for each semester. The most common graphs constructed were bar graphs and scatter plots, followed by line graphs, and lastly, box and whisker and dot plots. In the figure above, *n* indicates the number of total graphs created for that lab.

3.8.4 Qualitative Description of Graph Attributes

To address the second half of the third research question, we qualitatively describe the attributes of graphs constructed by undergraduate students. We use the previously established categories: graph mechanics, data form, graph choice, and aesthetics and visuo-spatial considerations (see Chapter 2, Appendix 2C), to describe students' graphs.

3.8.4.1 Graph Mechanics

Starting with graph mechanics, title, we noticed that across both semesters, and labs, the most common problem was that the title was not descriptive for the graph. Particularly, the title was missing the subject, details on the experiment, independent and dependent variables, and or was not in the form of a statement. We noticed a similar problem for axes labels. The labels were either incorrect or they were not descriptive for the type of data shown in the graph. For the x axis label, if the graph showcased categorical independent variables (e.g. bar graph, dot plot), a larger axis label was absent. For the y axis label, besides being descriptive for the experiment, if the data plotted were transformed values (e.g. averages, percentages, differences), this was not indicated on the y axis label. Units for each axes were usually accurate, however we did see some groups who forgot to include the units. Some common mistakes we noticed for the scale was the lack of axis breaks on either the x or the y axis, or inappropriate increments for their data.

3.8.4.2 Data Form

The next category, data form refers to the graph showing a clear distinction between raw and manipulated data. There were two common errors that were prevalent across two semesters, it was either difficult to discern between raw and manipulated data (See Figure 3.6) or if the group had plotted averages but failed to include standard deviation or standard error (See Figure 3.7).

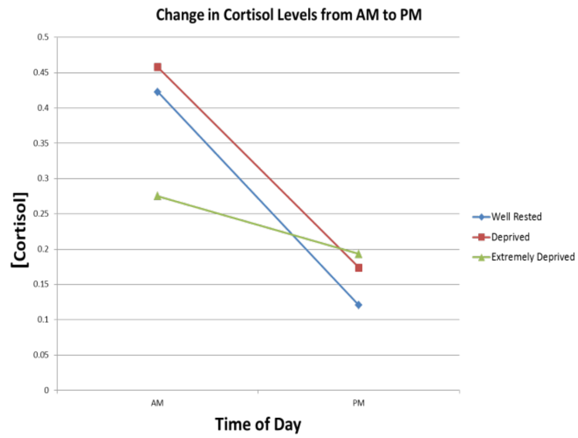


Figure 3.6: Difficult to discern between raw and transformed data values: This line graph shows how one group is trying to show a change in cortisol levels from AM to PM in three conditions of well rested, deprived, and extremely deprived individuals. However it is unclear if these data are average changes between multiple individuals in each condition or if the graph is showing a change in cortisol for one individual in each condition. One way to resolve this issue with data form is to include the sample size of individuals in each condition and to provide a more descriptive label on the y axis.

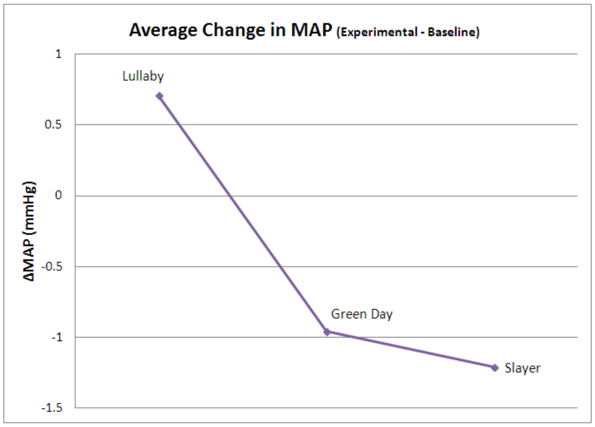


Figure 3.7: Lack of error bars: This line graph shows how one group plotted the average change in MAP, but did not include error bars to indicate variance. This was one of the common errors associated with data form in the Spring 2013 and 2014 semesters.

3.8.4.3 Graph Choice

For the third category, graph choice, when looking at graph type, usually the graphs were appropriate for the independent and dependent variables but there were instances where the data could have been grouped in a more effective way. Early on in the semester, when students collected data for earthworm action potentials, they generally plotted raw data in bar graphs (See Figure 3.5) because of the small sample sizes they collected. This is partly because students only had two hours to design their experiment, set up the nerve chamber, anesthetize their earthworms, familiarize themselves with the Lab Scope software to collect data, and troubleshoot problems that may have surfaced. However, by the end of the semester, students displayed transformed values and diversified their graph choice to include box and whisker and dot plots (See Figure 3.5). One reason for students plotting transformed values could be because students were given two lab periods, approximately six hours total to design their experiment and collect data. Furthermore, students were encouraged to participate in each other's' experiments, so many groups had a large dataset to work with. Another reason for students diversifying their graph choices is because of the suggestions made by AA and SMG to the students. Although we saw the most diversification of graphs and transformed data in the last lab, some groups during the second and third labs, chose to plot all raw data in series of bar graphs, where each graph showed data for one subject. This is an ineffective use of space and the data plotted did not align with the original intended purpose of the graph. Graphs displayed in Figure 3.8 are from one group and shows this finding.

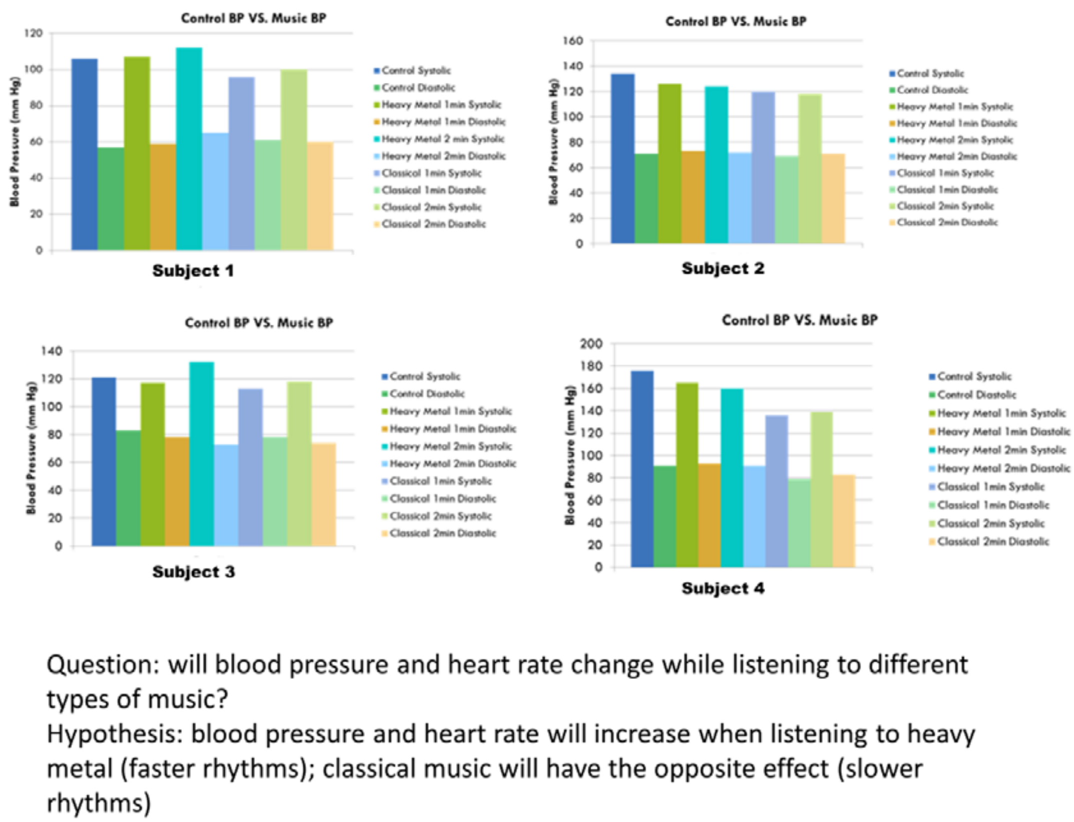


Figure 3.8: Ineffective use of data presentation by one group: These bar graphs show how one group made a separate graph for each experimental subject. Because the focus on individuals is not emphasized in the research question and hypothesis developed by the group, making separate graphs does not help the graph reader interpret the trends. A better solution to present these data would is to display them in one or two dot plots to show all of the raw data while communicating the variance in each group.

One of the reasons that a graph is constructed is so that it is easy for the graph constructor to convey findings from their experiment. Thus, in order to fulfill this purpose, a graph should be aligned with the research question and/or the hypothesis (College Board, 2012) For alignment, most times if the graphs were labeled appropriately, it was easy to see if there was an alignment or not. A graph was declared as partially aligned if either the research question or hypothesis or graph did not explicitly state information on the experiment or identify the subjects. The graphs in Figure 9 above are partially aligned to the research question and hypothesis because they did not specify that they were interested in reporting data for individual subjects or at particular time intervals.

Finally, graphs were evaluated for the quality of the take-home message. An appropriately constructed graph should allow the reader to draw appropriate conclusions from the data in the graph. However, lack of experimental details, descriptive labelling, or missing information on the subjects, made it difficult to formulate a take-home message.

3.8.4.4 Aesthetics and Visual-Spatial Considerations

The last category, aesthetics and visuo-spatial considerations states that a graph should be pleasing to the eye, the data plotted should occupy sufficient room in the Cartesian plane, and mechanistic properties should make it easy to extract meaning from the graph. If a group made multiple graphs for one experiment, it was common to see the same mistakes appear across different graphs. For example, the graphs in Figure 3.8 all display multiple bars of different colors. If the group wanted to display all treatments on one graph, they could have grouped and spaced the bars into the control, heavy metal, and classical music treatments. They could also have considered using a similar shade of color that would make it more aesthetically pleasing and easier to interpret trends.

3.9 Discussion

Graphing is one of the many important skills associated with science literacy (AAAS, 2011; Gormally et al., 2012). Providing students with ample opportunities in a classroom setting to work with data, practice skill development with graphing, and stop and reflect on their graphs is important for fostering critical thinkers, lifelong learners, and is a step forward in encouraging students to improve their MRC (Tynjala 1999; diSessa, 2004; McFarland, 2010). In order to assist a student at a novice or journeyman level to improve their MRC with graphing, targeted materials are necessary. However, prior to material development, it is important to gather information regarding students' basic knowledge with graphing. Although previous studies have documented student difficulties around graph choice and construction (Bray-Speth et al., 2010; McFarland, 2010), their students still exhibited difficulties and they did not investigate deeply into student reasoning, which was precisely the aim of this research.

In this study, we utilized student reflections to inform the development of graphing materials. Particularly, we wanted to see how students reasoned with graph choice, advantages, and disadvantages for their graphs, and what were the qualities of their graphs. We summarize our findings in terms of the categories of MRC below.

3.9.1. MRC Category, Functioning

Students talked a lot about *graph interpretation*, looking at comparisons, differences/change, relationship between data points/sets, statistical concepts, and the take-home message. Students also spoke about the *communication* aspects of the graphs such as figural effects, function of the graph, and graph components. This indicates that students can often link the usage and purpose of a graph to the type of data present, which is one of the main categories in MRC. We did not see students mention aligning their graph to their research question and hypothesis in the study. One reason could be that students are unaware that a graph should serve a specific purpose, usually declared by the research question and hypothesis. Incorporating alignment into our graphing learning materials, can further strengthen student's reasoning with the function of a graph.

3.9.2. MRC Categories, Critique and Learning or Reflection

Cursory analysis of student reflections with advantages and disadvantages highlight that students can often confidently and correctly identify the advantages, but they sometimes showed difficulty articulating the disadvantages of their graph. Common answers for the disadvantages of the graph were either incorrect, (e.g. if a student reflected on a scatter plot that showed all of their raw data, and they stated that a disadvantage was that “the graph doesn’t show detailed information”), or were actually advantages, (e.g. if a student reflected on a box and whisker plot, and they stated that they “cannot see the overall distribution of the dataset”). In both of these cases, students listed the purpose or advantage of either the scatter plot or box and whisker plot, instead of the disadvantages. Although one reason behind students difficulty with articulating the disadvantages could be due to the format in which information was collected (written instead of interviews), another reason could be that some students are unaware that graphs can have disadvantages. This could stem from the fact that students in classrooms are usually not asked to consider the tradeoffs and disadvantages of graphs they choose to represent their data with. The second way that we analyzed the advantages and disadvantages was through word cloud analysis. Although word clouds can quickly point out the major words from dense qualitative data, they do not automatically collapse across words with similar meanings, nor do they list the frequency of the words as they appear together. So our interpretation of these word clouds may not be accurate, since we are looking at single words. As part of MRC categories, critique and learning or reflection, we aim to incorporate into our lectures what constitutes a disadvantage, and how to write one. By doing this, we hope that when students are exposed to graphs in the media, they can critique the graph and assess if the graph is conveying misleading information or not.

3.9.3 MRC Category, Invention

Graphs were constructed by groups in the laboratory which makes it impossible for us to know if students took a reflective approach to graphing or if they clicked a few buttons on excel and created their graph. Therefore, it is difficult to reflect on the student reflection data within the context of the MRC category, invention. Although this is a limitation of this study, findings from our previous, expert-novice study, where think aloud interviews were conducted to understand stepwise reasoning with graph construction help to resolve this issue. Thus, one of the materials we aim to develop will allow students to practice and reflect on their graph construction, and allow them to think of other ways they can plot data, and assess the advantages and disadvantages of their graph representation. Also, some of the most common types of graphs constructed were either bar, scatter, or line graphs. In order to increase students' MRC with invention, they must be exposed to many different types of graphs, understand the function of each so that they know when it is appropriate to choose a particular type of graph over other options.

3.10 Project Scope and Future Studies

Four main study design features determine the scope of our conclusions. First, data were collected from 139 upper-level students at a single Midwestern United States research university, enrolled in an upper-division physiology course, with its unique curriculum and learning objectives for students. Furthermore, since we did not repeat this study in other laboratory classrooms with different student demographics, the claims presented here are only for the 139 undergraduate students enrolled in this course. To verify our findings fully, future work is needed at other types of institutions, in different disciplinary fields, and with their own undergraduate students in order to fully understand and appreciate what the reasoning is like for graph choice. McFarland (2010) did a similar activity in her classroom where students were asked to construct graphs by hand and evaluate their graphs. However, students did not actively engage in experimental design and were given a dataset. Furthermore, the graphing activity in McFarland (2010) was only conducted once and not multiple times over the semester.

Second, for the purpose of the purpose of the study, we did not focus on group dynamics or micromanage groups to ensure that every member of the group had the opportunity to construct graphs. However, we did announce for each new lab module that students should exchange group roles so that every person in the group gets the opportunity to engage in all the different roles (e.g. principle investigator, equipment specialist, data specialist, etc.). Since we did not keep track of students' roles in the group, it is quite possible that the person who felt the most comfortable using computer software constructed graphs for their group over the course of the entire semester. When students had questions about the types of graphs to construct, AA and SMG posed questions and involved the entire group so that they could all discuss the different types of graphs available to them. We hoped that with post presentation graph reflections, that all students would take the time to understand the reason behind their graph choice as a group, even if they were not the ones who constructed the graph. We suggest future studies can focus at the level of the student group to see how students engage in experimental design, data collection, analysis, and communicating their findings in the form of a graph. We also suggest encouraging each group member to identify a type of graph they would use for their graph and sketch a graph with the appropriate variables, prior to the groups' presentation. This will give everyone in the group an opportunity to think about the types of graphs available and would allow them to see if the graph they chose is appropriate for the type of data they have. Third, students were asked to reflect on their graphs for times over the semester, but we did not give them explicit instruction on how to engage in reflection, nor did we give the students feedback. We did not track the quality of reflections for each student over the course of the semester and neither did we associate the "correctness" of the graph with the graphing reflection Our reasoning was to see the types of things students said naturally, instead of biasing them towards specific vocabulary or terminology that they may not understand. Future studies should definitely consider presenting students with various examples showing different complexities of reflection and providing students with practice to reflect. Targeted feedback on students' reflections will help students with their metacognition, but may also positively impact the quality of their graph construction.

Fourth, student reflections were in written form, instead of oral, which may have been difficult for some students to articulate their reasoning for graph choice, advantages, and disadvantages. Conducting think-aloud interviews would have ensured that we obtained a thorough understanding of student reasoning with graph choice because the interviewer would have probed the participant to elaborate their thinking. However, we wanted to understand the thought behind graph choice in a naturalistic setting, and written reflections are best because they can be implemented in a short amount of time. Furthermore, the purpose of this study was to inform the development of instructional materials aimed at improving graphing (See chapter 4). We are confident in our coding scheme because we saw saturation of student responses within our secondary and primary codes, and themes, except for the theme *technology*, which was drawn out from our literature review. Interviews provide a rich source of data, however transcription, analysis, and interpretation may take a lot longer than coding written responses. Interviews are a great avenue for collecting data, but perhaps they can be implemented in a future study, with an external interviewer so that the teacher-student relationship is still maintained, otherwise students participating in interviews may feed "judged" afterwards.

3.11 Summary

In this chapter, we presented a study aimed at understanding the reasoning implemented by undergraduate students when choosing and creating appropriate graphical representations of physiological data, in an upper-division physiology laboratory course. Specifically, we wanted to understand the reasoning when students reflect on graph choice, advantages, disadvantages, and the types of graphs created and their attributes. Student reflections for graph choice fell into five main themes: technology, graph interpretation, communication, reflection, and experimental concepts, with the most common themes being graph interpretation and communication. There was no explicit mention of the question or hypothesis in the students' reflections on graph choice. Students often correctly identified the advantages, associating variable type with graph type. However, when articulating the disadvantages, students often stated the advantages or self-critiqued their graphs. The most common types of graphs constructed were: bar, scatter, or line graphs, and dot plots and box and whisker plots were the least common. With the graph attributes, we saw that students can improve by including descriptive graph mechanics, clearly distinguish between data form, improve their graph choice, alignment with the research question and hypothesis, and improve the aesthetics and visuo-spatial aspects of their graphs. Findings from this study will inform the development of graphing instructional materials that will target the aforementioned areas and improve student's approach and reasoning with graph choice and construction.

Acknowledgements

We would like to thank our research group, PIBERG, for their feedback on this project. We are indebted to all of the biology undergraduate students who participated in this study. This project emerged from ideas initiated within the Biology Scholars Research Residency program.

References

Åberg-Bengtsson, L., & Ottosson, T. (2006). What lies behind graphicacy? Relating students' results on a test of graphically represented quantitative information to formal academic achievement. *Journal of Research in Science Teaching*, *43*(1), 43-62.

Ainley, J. (2000). Transparency in graphs and graphing tasks: An iterative design process. The Journal of Mathematical Behavior, 19(3), 365-384.

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). Development of a framework for graph choice and construction. *Advances in physiology education*, *40*(1), 123-128.

Ates, S., & Stevens, J. T. (2003). Teaching line graphs to tenth grade students having different cognitive developmental levels by using two different instructional modules. *Research in Science & Technological Education*, *21*(1), 55-66.

Bertin, J. (1983). Semiology of graphics: diagrams, networks, maps. Madison, WI: Univ Wisconsin Press.

Bing, T. J., & Redish, E. F. (2012). Epistemic complexity and the journeyman-expert transition. *Physical Review Special Topics-Physics Education Research*, *8*(1), 010105.

Bodner, G. M. (2006). The role of theoretical frameworks in chemistry/science education. ln. Theoretical frameworks for research in chemistry/science education, 2-26J.

Bodner and Orgill (2007). The role of theoretical frameworks in chemistry/science education.

Brasell, H. M., & Rowe, M. B. (1993). Graphing skills among high school physics students. School Science and Mathematics, 93(2), 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*(3), 323-332.

Bruno, A., & Espinel, M. C. (2009). Construction and evaluation of histograms in teacher training. *International Journal of Mathematical Education in Science and Technology*, *40*(4), 473-493.

Bussey, T. J., Orgill, M., & Crippen, K. J. (2013). Variation theory: A theory of learning and a useful theoretical framework for chemical education research. Chemistry Education Research and Practice, 14(1), 9-22

College Board. (2012). AP Biology Quantitative Skills: A Guide for Teachers. Retrieved from http://apcentral.collegeboard.com/apc/public/repository/AP_Bio_Quantitative_Skills_Guide-2012.pdf

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*(3), 317-322.

Cooperstein, S. E., & Kocevar-Weidinger, E. (2004). Beyond active learning: a constructivist approach to learning. *Reference Services Review*, *32*(2), 141-148.

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

DiFazio, M. H. (1990). Graphics Software Side by Side. The Mathematics Teacher, 83(6), 436-446.

diSessa, A.A., & Sherin, B. L. (2000). Meta-representation: An introduction. *The Journal of Mathematical Behavior*, *19*(4), 385-398.

diSessa, A. A. (2004). Metarepresentation: Native competence and targets for instruction. *Cognition and instruction*, *22*(3), 293-331.

Duffy, T. M., & Jonassen, D. H. (1992). Constructivism: New implications for instructional technology. Constructivism and the technology of instruction: A conversation, 1-16.

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*(22), 3040-3059.

Elliott, A.C., Hynan L.S., Reisch, J.s., & Smith, J.P. (2006). Preparing Data for Analysis Using Microsoft Excel. *Journal of Investigative Medicine*, 54(6), 334-341.

Few, S. (2004). *Show me the numbers: Designing tables and graphs to enlighten* (Vol. 1, No. 1). Oakland, CA: Analytics Press.

Fosnot, C. T. (86). Perry, RS.(2005). Constructivism: a psychological theory of learning. CT Fosnot (ed).

Friel, S. N., & Bright, G. W. (1996). Building a Theory of Graphicacy: How Do Students Read Graphs?. Paper Presented at the Annual Meeting of AERA, New York.

Gormally, C., Brickman, P., Hallar, B., & Armstrong, N. (2009). Effects of inquiry-based learning on students' science literacy skills and confidence.*International journal for the scholarship of teaching and learning, 3*(2), 16.

Gormally, C., Brickman, P., & Lutz, M. (2012). Developing a Test of Scientific Literacy Skills (TOSLS): measuring undergraduates' evaluation of scientific information and arguments. *CBE-Life Sciences Education, 11*(4), 364-377.

Gross, L. J. (2004). Points of view: the interface of mathematics and biology interdisciplinarity and the undergraduate biology curriculum: finding a balance. *Cell Biology Education*, *3*(2), 85-87.

Hattikudur, S., Prather, R. W., Asquith, P., Alibali, M. W., Knuth, E. J., & Nathan, M. (2012). Constructing Graphical Representations: Middle Schoolers' Intuitions and Developing Knowledge About Slope and Y-intercept. *School science and mathematics*, *112*(4), 230-240.

Kanari, Z., & Millar, R. (2004). Reasoning from data: How students collect and interpret data in science investigations. *Journal of Research in Science Teaching*, *41*(7), 748-769.

Konold, C., Higgins, T., Russell, S. J., & Khalil, K. (2015). Data seen through different lenses. *Educational Studies in Mathematics*, *88*(3), 305-325.

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

Labov, J. B., Reid, A. H., & Yamamoto, K. R. (2010). Integrated biology and undergraduate science education: a new biology education for the twenty-first century?. *CBE-Life Sciences Education*, *9*(1), 10-16.

Leonard, J. G., & Patterson, T. F. (2004). Simple computer graphing assignment becomes a lesson in critical thinking. NACTA journal, 48(2), 17-21.

May, D. B., & Etkina, E. (2002). College physics students' epistemological self-reflection and its relationship to conceptual learning. *American Journal of Physics*, *70*(12), 1249-1258.

McFarland, J. (2010). Teaching and Assessing Graphing Using Active Learning. *MathAMATYC Educator*, *1*(2), 32-39.

McNaught, C., & Lam, P. (2010). Using Wordle as a supplementary research tool. *The qualitative report*, *15*(3), 630.

Millar, N. (2001). Biology statistics made simple using Excel. *School Science Review*, *83*, 23-34.

Ohland, M. W., Loughry, M. L., Carter, R. L., Bullard, L. G., Felder, R. M., Finelli, C. J., …& Schmucker, D. G. (2006, June). The comprehensive assessment of team member effectiveness (CATME): A new peer evaluation instrument. In Proceedings of the 2006 ASEE Annual Conference.

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*(1), 20-26.

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*(1), 95-115.

Patterson Jr, T. F., & Leonard, J. G. (2007) E-Graphing: Perfection, Paradox or Persuasion. http://www.uvm.edu/~jpetrill/Paterson%20&%20Leonard%20-%20E-Graphing.pdf. Accessed August 13, 2016.

Patton, M. Q. (2005). *Qualitative research*. John Wiley & Sons, Ltd.

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.

Saldaña, J. (2015). The coding manual for qualitative researchers. Thousand Oaks, CA: Sage.

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*(1), 75-87.

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*(4), 690.

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

Tairab, H. H., & Khalaf Al-Naqbi, A. K. (2004). How do secondary school science students interpret and construct scientific graphs?. *Journal of Biological Education*, *38*(3), 127-132.

Trolier, T. K., & Hamilton, D. L. (1986). Variables influencing judgments of correlational relations. *Journal of Personality and Social Psychology*, *50*(5), 879.

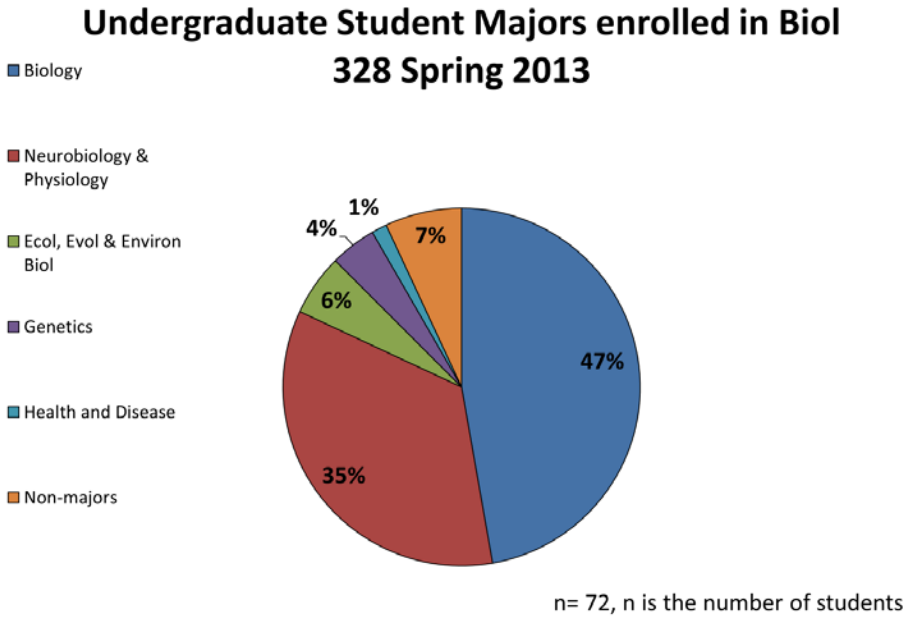
Tufte, E. R., & Graves-Morris, P. R. (1983). *The visual display of quantitative information* (Vol. 2, No. 9). Cheshire, CT: Graphics press.

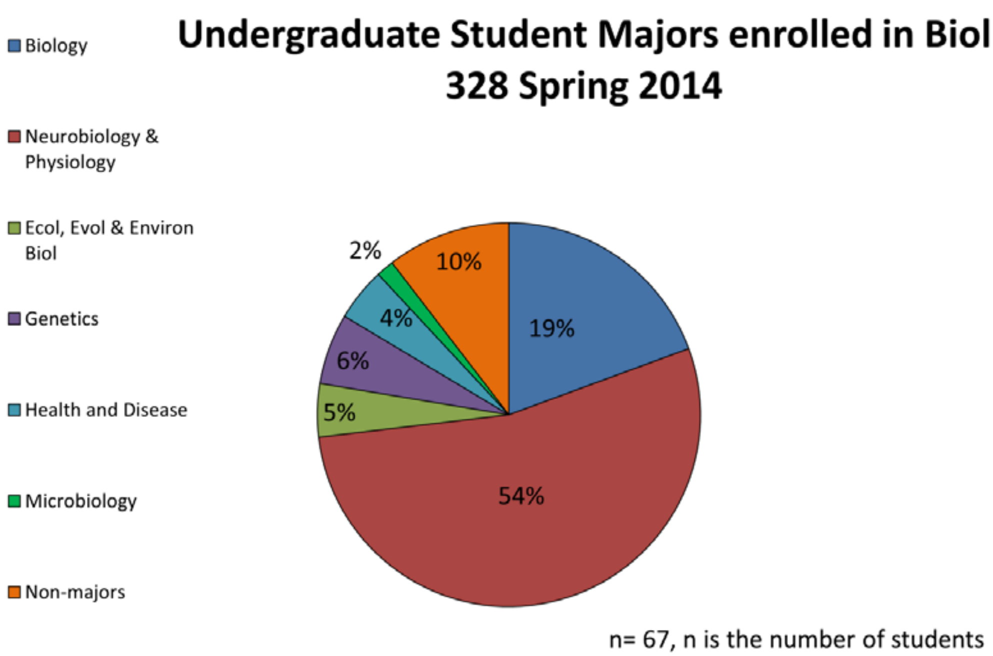
Tynjälä, P. (1999). Towards expert knowledge? A comparison between a constructivist and a traditional learning environment in the university. International journal of educational research, 31(5), 357-442.

Webber, H., Nelson, S. J., Weatherbee, R., Zoellick, B., Schauffler, M.(2014). The Graph Choice Chart. *The Science Teacher 81*, 37-43.

WordClouds.com. retrieved from http://www.wordclouds.com/ August 17, 2016.

Appendix 3A: Demographics of Undergraduate Students Enrolled in Biol 328 in Spring 2013 and 2014 Semesters





Appendix 3B: A Guide to Graphs and Tables

The graphic presentation of data is a very important aspect of communicating ideas in science. Whatever the format of presentation, (slide-show, poster, abstract, or manuscript) your audience often makes critical judgments based on the data you present them with. As such, learning to effectively present the data you have collected will help you tremendously in the expression of your own ideas. This handout is designed to reacquaint you with scientific graphing.

When assembling graphs for presentation, be careful not to include all accessible data. Presenting all material collected typically obscures the major point of your manuscript. To avoid this, before you assemble the outline of your presentation carefully establish the major points you'd like to make. For example, you have measured the resting heart rate in beats per minutes (bpm) of everyone in your lab group. You calculate the mean value to be 71.32 bpm. How should you report it? 71 bpm?, 71.3 bpm? or 71.32 bpm? You need to consider what you are measuring and the sensitivity of your instruments. In this case, reporting a fraction of a heartbeat in the average doesn't make sense because you don't have a third of a heart beat! If multiple trials were performed in a given condition, the average of the trials with the standard deviation should be plotted for each condition.

Below are examples of the most common graphical displays of data for your reference. Each of the computers in the lab has the PowerLab data acquisition and analysis software. Within 'Chart' you should be able to get all the measurements that you are interested in. I would suggest using Microsoft Excel or OriginPro on the laptops to generate figures for data other than example traces of recordings. Appropriate statistics can also be performed within Excel and OriginPro. Please consult with me regarding the appropriate statistical tests to perform on your data and how to display those results on your figures.

Raw data plots

An extremely useful type of graph that can be used either for one of your data displays to report with or to get an idea of the data distribution to inform further analysis decisions is the histogram. The measured values are plotted on the x-axis and the number of times that value is observed is plotted on the y-axis. Often the x-axis values are organized into ranges called bins. Below are histograms of nerve conduction velocities in different nerves (Figure 1)

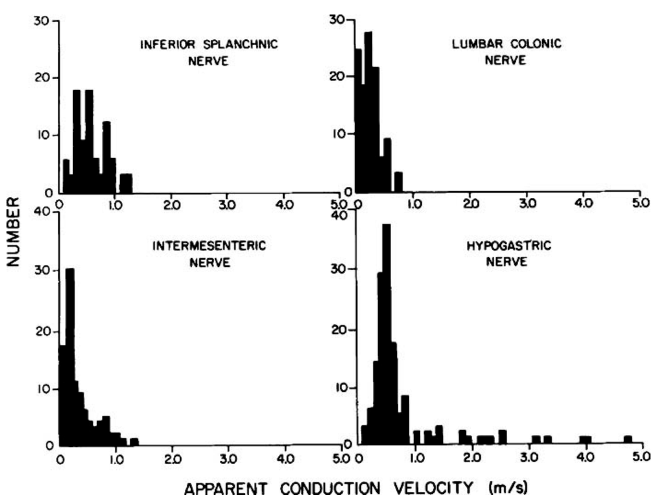


Figure 1 - Example of a histogram

Figure 11. Frequency histogram of conduction velocities of preganglionic fibers in different trunks attached to nerve inferior mesenteric ganglion of guinea pig.
(Source: Data from J. H. Szurszewski and P. J. Crowcroft.)

Appendix 3B continued

Categorical data sets

Bar graphs

Bar graphs are useful in comparing (or contrasting) a series of values collected under different circumstances that cannot be conveniently arranged in a quantitative manner. They are also useful in comparing a series of values collected at the same time. For example, the data shown in Figure 2 below are from an experiment evaluating the time it took for subjects to exert themselves to exhaustion (y-axis, **dependent variable**) depending on prior activity or recovery activities (x-axis categories, **independent variables**). Note that the x and y axes are clearly labeled, each group is easily distinguishable, and the standard error is shown as a line on each bar. Significant differences are indicated with asterisks. Also note the figure legend under the figure describing in text what is plotted with some more information to aid in its interpretation.

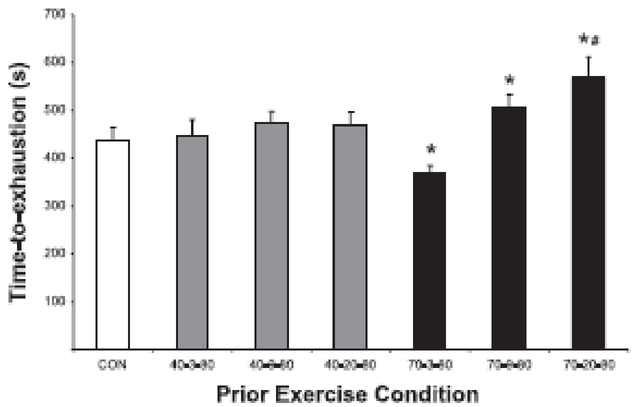


Fig. 6. Mean ± SE times to exhaustion during severe-intensity exercise in control and after the various prior exercise and recovery permutations. Relative to control, the time to exhaustion was increased by 2% in the 40-3-80 condition [nonsignificant difference (NSD)], by 8% in the 40-9-80 condition (NSD), by 7% in the 40-20-80 condition (NSD), by 15% in the 70-9-80 ($P < 0.05$), and by 30% in the 70-20-80 condition ($P < 0.05$). The time to exhaustion in the 70-20-80 condition was significantly greater than in the 70-9-80 condition ($P < 0.05$). Relative to control, the time-to-exhaustion was 16% lower in the 70-3-80 condition ($P < 0.05$).

Figure 2 - Example of a bar graph (From Bailey et al., 2009)

Dot plots

Dot plots are a great way to show all of the data in a data set so the viewer can see the distribution easily. The mean is often denoted with a bar. Figure 3. is an example of a dot plot.

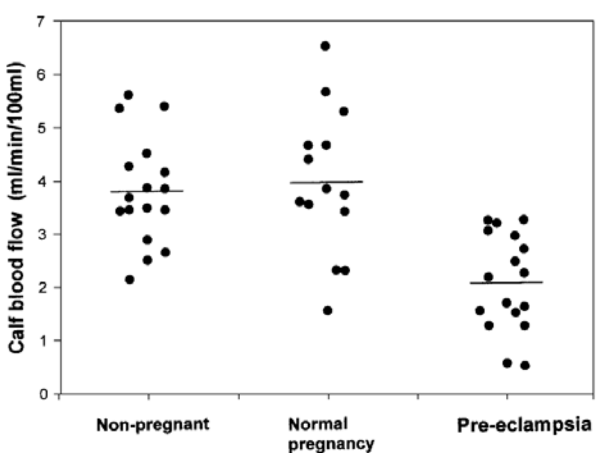


Figure 3 - Example of a dot plot (From N. ANIM-NYAME et al., 2000)

Figure 1 Dot plot comparing calf nutritional blood flow in non-pregnant controls, normally pregnant controls and women with pre-eclampsia

Appendix 3B continued

Box plots

Box plots are a statistical plot that allows you to provide very detailed descriptive statistics about a data set in a concise visual. The median of the data set (line in the middle of the box) is determined and then the data set is divided into four, equal parts (quartiles) along the range of the data values. The number of data points that fall into each of those quartiles is represented by the size of the box (for the two middle quartiles around the median) or whisker for the outer quartiles. The mean is often denoted with an asterisk or other symbol.

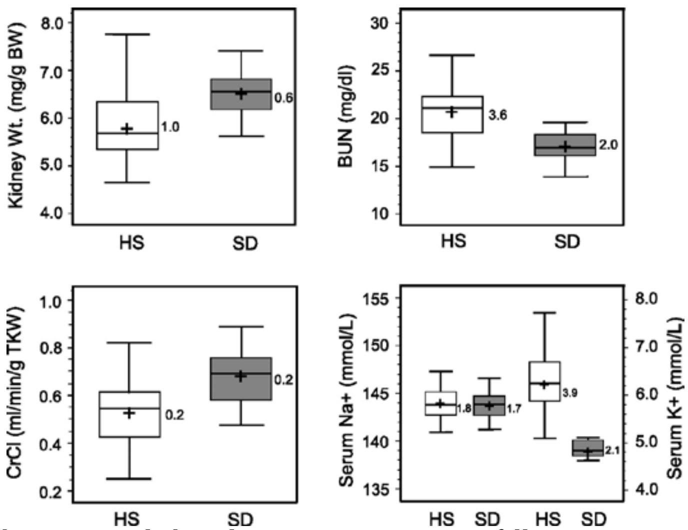


Figure 4. Example of a box plot (From: http://ajprenal.physiology.org/cgi/content-nw/full/298/6/F1484/F2)

Fig. 2. Box plot comparing physical traits and renal parameters of HS and Sprague-Dawley (SD) rats. Open plots are HS data, and shaded plots are SD data. The box itself contains the middle 50% of the data. The upper quartile (UQ) indicates the 75th percentile of the data set, and the lower quartile (LQ) represents the 25th percentile. The interquartile range (IQR = UQ – LQ) represents a measure of variability and is displayed on *right* of each box plot. The line within the box indicates the median value of the data; + represents the mean value of the population. BW, body weight; BUN, blood urea nitrogen; CrCl, creatinine clearance.

Appendix 3B continued

Continuous data sets

Line Graphs

Line graphs are most often used to express how one measured variable varies continuously with another variable. In general one variable such saturation of hemoglobin with O_2 will depend on another variable such as the amount of O_2 available (P_{O2} , the Figure 5). The hemoglobin saturation in this example is the **dependent variable** and P_{O2} is the **independent variable**. Convention favors plotting the independent variable along the x-axis or abscissa, and the dependent variable along the y-axis, or ordinate.

To distinguish between the different groups to be plotted, it is recommended that obviously different symbols be used for each group or that they be clearly labeled if space permits you to do so clearly. Below are two examples of line graphs. The figure of left shows the hemoglobin saturation vs. PO2 for in different pH (top) and in different birds (bottom). The figure on the right is a plot of several dependent variables measured during a dive by an emperor penguin as a function of time. Note how the different groups and measure are plotted in a way that makes them easily distinguishable from one another different colors and/or symbols).

Figure 5. Examples of line graphs. (From: Meir and Ponganis, 2009).

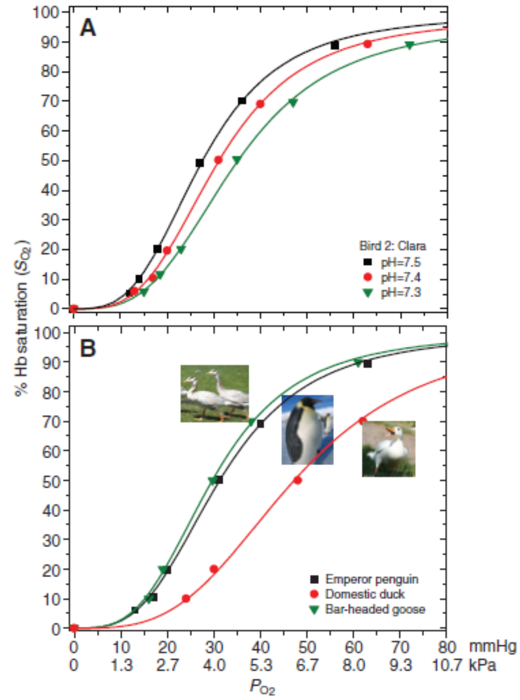


Fig. 1. An example of oxygen–hemoglobin (O_2–Hb) dissociation curves from (A) one penguin at pH 7.5, 7.4 and 7.3, and (B) the emperor penguin, the bar-headed goose (*Anser indicus*) (Black and Tenney, 1980) and the domestic duck (*Anas platyrhynchos*, forma domestica) (Hudson and Jones, 1986) at pH 7.4. Note that as for the bar-headed goose, the O_2–Hb dissociation curve of the emperor penguin is significantly left-shifted as compared with the domestic duck (and most birds). The bar-headed goose photo is courtesy of Graham Scott; the domestic duck photo is by Maren Winter (licensed under the terms of the GNU Free Documentation License, Version 1.2 or any later version); the penguin photo is by J.M.

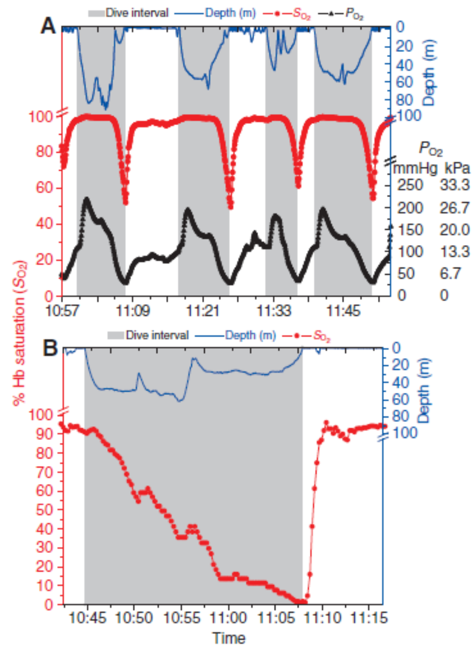


Fig. 2. The S_{O_2} profile during (A) 1 h of diving of bird 1 (2008) (arterial S_{O_2}) with P_{a,O_2} superimposed. Note that $P_{a,O_2}/S_{a,O_2}$ are at low levels at the start of the measurements because a dive had occurred just before the series recorded, and (B) the current record dive (23.1 min) of an emperor penguin [bird 19 (Ponganis et al., 2007), venous S_{O_2}]. Note the arterialization of the venous blood O_2 store in this dive, as S_{v,O_2} before the dive is as high as 95% and the initial S_{v,O_2} of the dive is 91%. S_{v,O_2} decreased to 1% by the end of this long dive. S_{O_2} was determined at pH 7.5 throughout the entire dive to maintain consistency and to provide a conservative estimate of continuous S_{O_2}.

Other data displays

Scatter plots:

Scatter plots are a nice way to illustrate how two variables are related to one another. The individual, raw data are plotted for the two values instead of averages. In the example below, the levels of sodium in the blood serum (liquid portion) are plotted as a function of the change in body weight following an endurance race. Also included are clinically-related classifications for sodium levels and the hydration status of the athletes. You can see that there is a relationship between the two variables.

Figure 6. Example of a scatter plot. (From: Rosner and Kriven, 2007)

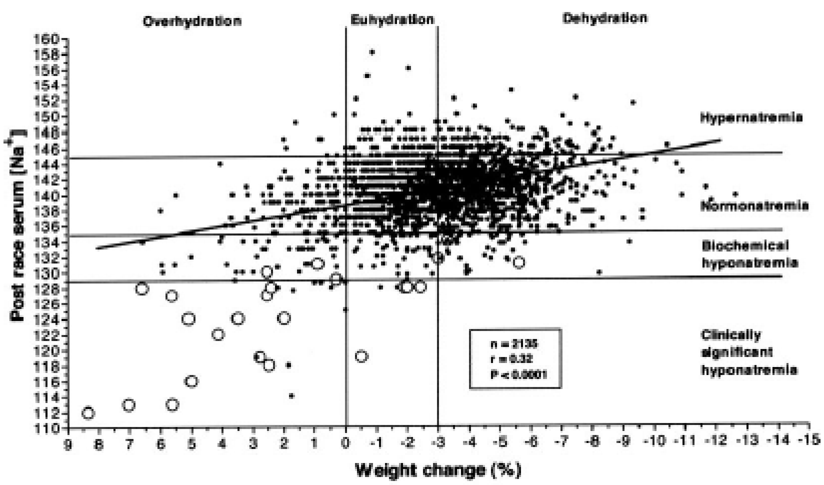


Figure 2. Relationship between serum sodium after racing and the weight change (in %) during exercise in 2135 athletes who competed in endurance events. ●, asymptomatic athletes; ○, athletes with symptoms compatible with EAH encephalopathy (EAHE). The majority of athletes who develop clinically significant hyponatremia have positive weight changes. Reprinted from reference (34), with permission. Copyright 2005 National Academy of Sciences.

Appendix 3B continued

Tables:

Tables are used to describe large amounts of data that would ordinarily be too burdensome to graph. The format of a table allows many variables which may have distinctly different units of measure to be assembled for quick reading. The table below is an example of a table summarizing heart rate, blood pressure, and exercise duration and intensity in men and women. Note the descriptive title at the top and the caption below providing some more information to assist in evaluating the data in the table.

Figure 7. Example of a table (From Chalela et al., 2009)

Table 4. Treadmill electrocardiographic exercise stress test characteristics of the men and women of the study sample.

Variables	Men		Women	
	N	Mean ± SD	N	Mean ± SD
Baseline heart rate (bpm)	200	77.9 ± 0.9	241	82.9 ± 0.8*
Peak exercise heart rate (bpm)	200	171.5 ± 1	241	170.1 ± 0.9
Heart rate recovery 1st min (bpm)	155	143.5 ± 1.1	207	139.2 ± 1.3*
Heart rate recovery 2nd min (bpm)	180	123.2 ± 1.1	216	117.2 ± 1.2*
Baseline systolic blood pressure (mmHg)	200	124.5 ± 0.7	241	119.5 ± 0.7*
Peak exercise systolic blood pressure (mmHg)	200	181.0 ± 1.6	241	166.8 ± 1.7*
Baseline diastolic blood pressure (mmHg)	200	78.7 ± 0.6	241	77.1 ± 0.6
Peak exercise diastolic blood pressure (mmHg)	200	84.2 ± 0.9	241	81.5 ± 0.8*
Rate-pressure product (bpm x mmHg)	200	31023.6 ± 321.7	241	28378.7 ± 313.1*
Exercise duration (min)	200	8.4 ± 0.1	241	6.9 ± 0.1*
Exercise capacity (MET)	188	12.7 ± 0.2	228	10.4 ± 0.2*

MET = metabolic equivalent. *$P < 0.05$ compared to men (Student *t*-test).

Appendix 3C: Lab Figure Presentations (10pts each)

At the end of most experimental lab days (or the start of the next lab period, if time is short), each group will present a brief PowerPoint presentation to the class which summarizes the findings from an activity in lab. This will preferably be for data from your independent experiment, but if there are none then data from the prescribed activities are appropriate to present. Please refer to the Guide to Figures and Graphs handout on Blackboard to help you decide on the appropriate way to best display your data and consult with me for any inferential statistics you could perform.

In an effort to facilitate your communication of the data you collect, I would like you to use PowerPoint to construct ~6 slides to present to the class. These slides will explicitly cover the information expected from your figure presentations (see below), much of which is taken from the worksheet for each day. These slides should contain bullet points that you expand upon.

Slides 1-2 – Introduction slide(s)

- Background information and initial observation
  - This is the place for you to make connections with what we are learning in lecture and I encourage you to use slides/images from lecture here (with the source cited).
- Question, hypothesis, and predictions

Slide 3 – Method slide

- Control and test conditions, as appropriate
- Statements about what is being measured and how

Slides 4-5 – Figure slide(s)

- The figure title should state the take home message of your study

Slide 6 – Conclusions, limitations and future direction

- Take home message and statement about whether your data support your hypothesis.
- Sometimes you don't realize the limitations in your experimental approach until you actually perform the experiment and analyze the data. State these, if needed.
- What would you do next?

Appendix 3C continued

You will be assessed according to the following evaluation scheme where:
√ = 1 pt outcome met
- = 0.5 pts outcome partially met
o = 0 pts outcome not met or not present

Metric	**Outcomes**				**Score**
Figure(s)	√ □ - □ o □ Appropriate display	√ □ - □ o □ Labeled axes and title	√ □ - □ o □ Descriptive and inferential statistics	√ □ - □ o □ Visually accessible	/4
Presentation	√ □ - □ o □ Rationale, hypothesis, and prediction	√ □ - □ o □ Basic experimental approach	√ □ - □ o □ Description of the figure and analysis	√ □ - □ o □ Conclusions	/4
Oral Communication	√ □ - □ o □ Appropriate detail	√ □ - □ o □ Clarity			/2
Total					/10

1. Figure(s)
    a. Appropriate display
        i. Use of a bar graph, scatter plot, or line graph appropriate to:
            1. The type of data plotted
            2. Your hypothesis
    b. Labeled axes and title
        i. Descriptive title for the graph that summarizes the main finding and axes that are clearly labeled with units
    c. Descriptive and inferential statistics
        i. Plots of averages with standard deviations (descriptive statistics) are preferred to single observation data points. If appropriate, control data should be plotted on the same graph and any results of inferential statistics (ex. t test) should be noted on the figure (ex. An asterisk over data points that are statistically different from control values).
    d. Visually accessible
        i. Is the main finding clear by eye to the observer? (graphs not too busy)
        ii. Are the axes labeled and readable (font size)?
        iii. Are all symbols/color schemes defined?
2. Presentation
    a. Rationale, hypothesis, and prediction
        i. The rationale for performing the experiment (stating the background information and/or previous observation and what the question is)
        ii. Clear statement of the hypothesis to be tested
        iii. Predictions should be stated
    b. Basic experimental approach

ii. Presentation indicates prior planning by the group about who is saying what

Appendix 3C continued

i. A very brief overview of what was manipulated and measured in the experiment

c. Description of figure (TAKE YOUR TIME AND WALK US THROUGH YOUR DATA)
    i. A walk-through of the figure saying what is plotted as a function of what
    ii. Take the time to present the analysis of your data here
        1. Different plots stress different things in the data, so be sure to communicate this
    iii. Statement of the conclusion/take-home message from the plot
d. Conclusions
    i. Do the data support your hypothesis?
    ii. Limitations
        1. Technical or practical?
        2. What were the variables that you couldn't control? Where there any unexpected variables?
    iii. What would you do next?

3. Oral Communication
    a. Appropriate detail (I will give you a lot of feedback on the first presentation)
    b. Clarity
        i. Information communicated clearly

Appendix 3D: Themes, Primary and Secondary Codes for Graph Choice

Themes	**Primary Codes**	**Secondary Codes**	**Sample Quotes/Buzzwords**
Technology *Any mention of Technology used to construct the graph* References: Leonard and Patterson 2004 Patterson and Leonard 2005 Patterson and Leonard 2007	Excel		"Excel" "This was the only type of graph I knew how to make in Excel"
Graph Interpretation *"Proper graph interpretation includes: being able to define the relationship between variables; ability to reasonably interpolate and extrapolate; compare and relate the results of a given graph to natural phenomenon in question." (Padilla, McKenzie, Shaw, 1986)* Buzzwords: trend, trendline, relationship, change, difference References: Millar 2001- relationship Ainley 1995- intuitive approaches taken by young children (ages 8-10) Ates & Stevens,2003- relationship between variables Brassel & Rowe 1993 Bruno & Espinel 2009 Page 177 Dori & Sasson 2006 ; McKenzie and Padilla 1986; Freedman & Shah(relationships)	Comparison	Between Subjects	"compare trials between the individual subjects."
		Within Subjects	"compare trials within each subject."
		Between Treatments	"comparison of data obtained from subjects to compare ventilation rate for fast and slow songs"
		Between Data	"best form to allow for comparison among our data."
		Between Measured Variables Elliot et al 2006-Table 1 Bruno & Espinel 2009 Page 476 Trolier and Hamilton 1986 Tairab & Al Naqbi, 2010 Shah, Meyer, Hegarty 1999	"easiest visual representation comparing the independent variable (treatments) on the x-axis vs. the dependent variable of average tidal volume and average Δ tidal volume on the y-axis."
		Personal Opinion Buzzwords: Easy, best, clear, etc.	"easy to compare and apprehend the three breathing parameters (ERV, IRV and VC) measured in liters."
	Difference/Change	Between Subjects	"it showed the differences in gender in concern to each body position for the variable of vital capacity"
		Within Subjects	"Bar graph to explain the differences between heart rate change for each subject each of the two trials."
		Between Treatments	"wanted to compare the differences of before and after breath holding."
		Between Data	
		Between Measured Variables Elliot et al 2006-Table 1 Bruno & Espinel 2009 Page 476 Trolier and Hamilton 1986 Tairab & Al Naqbi, 2010 Shah, Meyer, Hegarty 1999	"different colored bars showed the different between lying and standing positions, again broken down for each height." "represent our data for changes in blood pressure"
		Personal Opinion Buzzwords: Easy, best, clear, etc.	"easier to visualize the difference among each parameter we studied."
	Relationship	Between Data	
		Between Measured Variables Elliot et al 2006-Table 1 Bruno & Espinel 2009 Page 476 Trolier and Hamilton 1986 Tairab & Al Naqbi, 2010 Padilla McKenzie, Shaw 1986 Shah, Meyer, Hegarty 1999	"best way to show the relationship between the categorical variables(each subject), and the quantitative variable on the y-axis."
		Personal Opinion Buzzwords: Easy, best, clear, etc.	"We chose bar graph for presenting cortisol level and mood. 1. Mood is treated as discrete variables so three categories. 2. Cortisol level is continuous so bar graph gives a good presentation of this relationship."
	Statistical Concepts Millar 2001 Trolier and Hamilton 1986 Hattikudur et al 2012 Picone 2007	Descriptive	"change in average tidal volume for all subject (between treatment)."
		Inferential	"A paired T-test was used because we were comparing before and after conditions for the same set of subject."
	Take-home message	Shah, Meyer, Hegarty 1999	"baseline and test condition results on the same graph to clearly visualize the effects of the increased dead space on TV and RR."

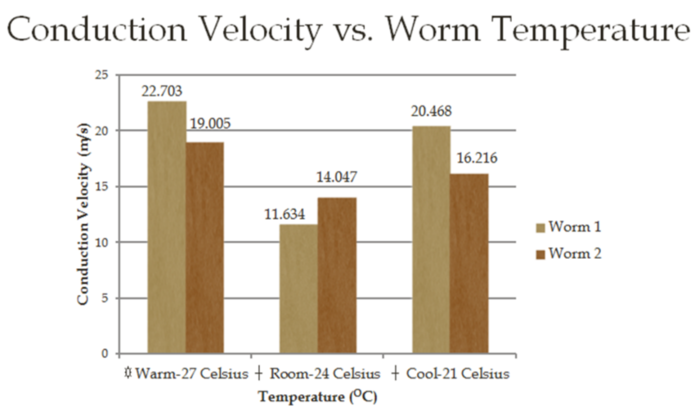
Appendix 3D continued

Themes	**Primary Codes**	**Secondary Codes**	**Sample Quotes/Buzzwords**
<u>Communication</u> *Consists of commenting on the appearance, aesthetics, graph quality, and graph functions.* <u>Buzzwords:</u> shows, represented, communicated, etc.	Figural Effects (DiSessa 2004)	Aesthetics Buzzwords: Color, synonyms and antonyms of neatness, “crowded”, “busy”	“Colored bars” “different colored bars showed the different between lying and standing positions, again broken down for each height.”
		Gestalt Buzzwords: Referring to the patterns, organization of data, parts of a whole	“the difference in colors and in the height of the bars.”
		Quality of Graph (Personal Opinion	“accurately display a large amount of data in a clear way.” “best way to show this was the same as the first graphs but only the tested variation for the hypothesis.”
	Function Describes the purpose and usage of the graph (Angra & Gardner, 2016)	Data	“The bar graph was able to illustrate our numerical and categorical data in a side by side chart while showing the difference between the two data sets.”
		Statistical Concepts	“We used a bar graph to show the averages over the 3 days for the different qualities of sleep”
	Graph Components Bengtsson 2006 Schriger and Cooper, 2002 Cooper et al 2001 Kosslyn 1994 DiFazio 1990 Elliot et al 2006 Leonard and Patterson 2004 Patterson and Leonard 2007 Ainley 2000 Brasell&Rowe 2003	Variables Buzzwords: mention of parts of a graph: axes, scales, key, what was plotted data points, etc.	“Thus, we made distinct taxonomies for caffeine, none-caffeine, am, pm + used a linear scale for cortisol levels.”
		Describing the layout of the entire graph Buzzwords: columns, bars, represent X, etc.	“The bar graph easily exhibited our averages(delta) for the differing scale choices over the three days.”
		Linking the type of data to graph type.	“bar itself shows the average.”

Appendix 3D continued

Themes	Primary Codes	Secondary Codes	Sample Quotes/Buzzwords
Reflection	Suggestions for Improvements *Buzzwords: Things that would have been done or could be done in the future.*	"bar graph could have effectively shown the differences between before and after rates, but would have been cluttered when trying to mash in trials and subjects."	
	Other Graphs *Buzzwords: any mention of other graphical representations than the one constructed.* (Schriger and Cooper, 2002) Published this paper to guide authors when making graphs and for editors to evaluate the graph quality (Table 1; Table 3 for advantages, disadvantaged) (Humphrey et al, 2013) Definitions between bar graphs and histograms (Franzblau and Chung, 2012) Advantages and disadvantages(Table (Kosslyn 1994) Book chapters 2,4,5,6	Generic	"Other graphical options did not allow us to present the data in the most effective manner."
		Other Graph Types	"This is more advantageous than a line graph because we're looking at 2 distinct time points rather than a timeline of measurements."
		Aesthetics	"Other types of graphs, such as a bar graph, would not be pro…tional, as there would be a large number of bars (one for each individual)"
		Data Buzzwords: Describing data as qualitative,quantitative, categorical, continuous	"categorical data would have been all stacked on top of each other for each group."
		Variables Elliot et al 2006-Table 1 Bruno & Espinel 2009 Page 476 Trolier and Hamilton 1986 Tairab & Al Naqbi, 2010 Shah, Meyer, Hegarty 1999	"we decided that temperature would be better/more fairly represented as a categorical variable"
Experimental Concepts <u>Buzzwords:</u> talking about experimental details related to purpose, hypothesis, methodology, subjects, treatments, trials, limitations, etc.	Research Question/Purpose	Schriger and Cooper, 2002; Franzblau and Chung, 2012; Kosslyn 1994 page 21 Patterson and Leonard 2005 Ainley 2000 Wainer 1984	"we were interested in displaying the change in heart rate, blood pressure, and MAP of each person"
	Hypothesis	"We used this to support our hypothesis that women have higher average cortisol levels."	
	Data Collection/Manipulation	"show data taken at certain points if time rather than data taken over time intervals." "We all did four 3 trials"	
	Limitations	Time	"because of time constraints + lack of a precise way to measure temperature."
		Samples	"only had 4 subjects" "Our data set was very limited because of the squirmy problems our worm caused during the experiment."
	Variables Elliot et al 2006-Table 1 Bruno & Espinel 2009 Page 476 Trolier and Hamilton 1986 Tairab & Al Naqbi, 2010	"visualize the data we collected for the control and compare it to the experimental variable data across the subjects." "it made sense to use a bar graph because we had more than one independent variable (cold temp 21 degrees C, room temp 24 degrees C, and warm temp 27 degrees C)"	
	Personal Opinion Buzzwords: Easy, best, clear, etc.	N/A	

Appendix 3E: Example of Coding



Technology

Graph Interpretation

Communication

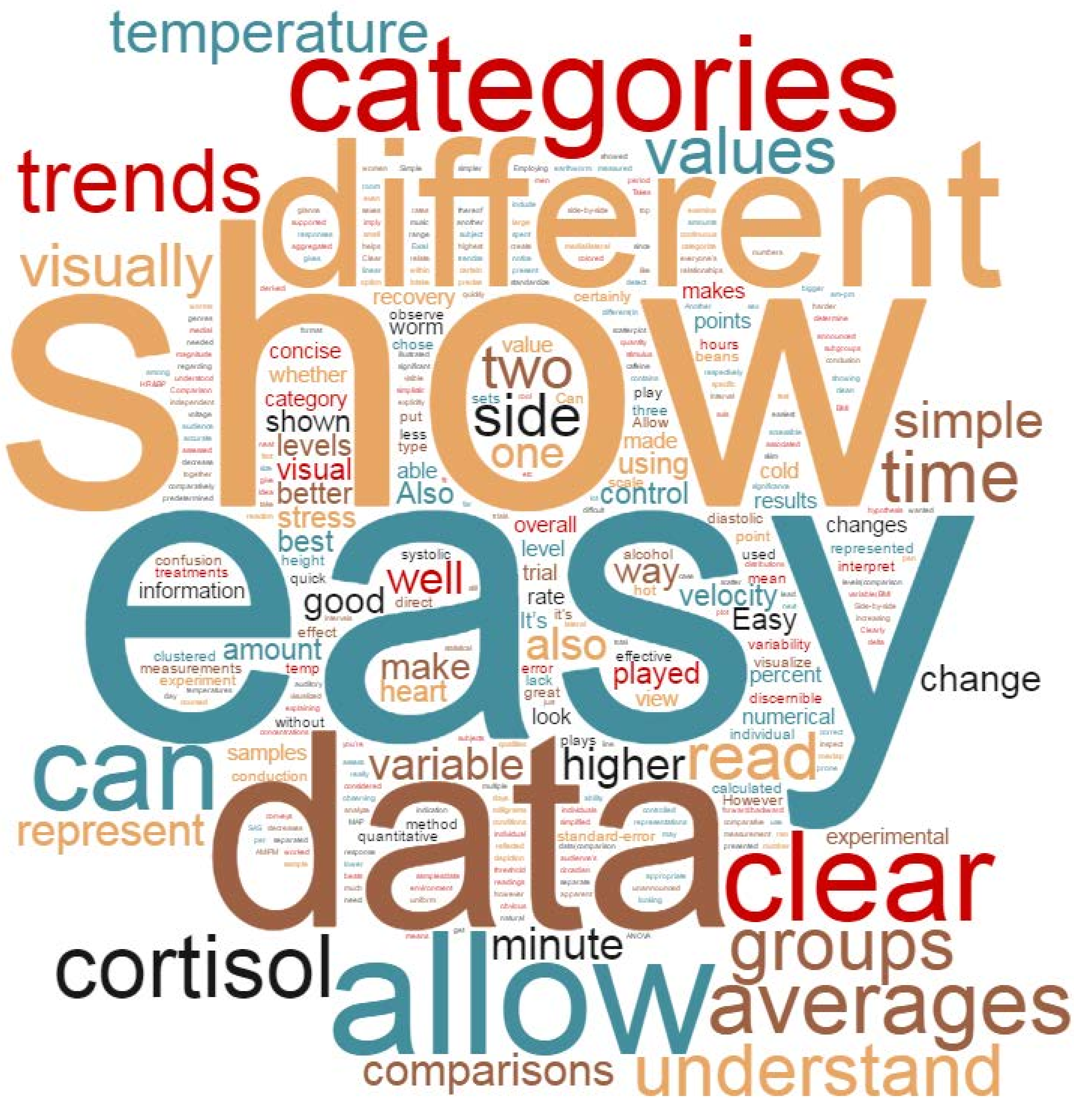
Reflection

Experimental Concepts

A3_P1 The categorical bar graph was chosen because it best summarizes the change of conduction velocity at different temperature condition and it shows clear comparison between velocities. And the temperature of worm for each condition (warm, room temperature, cool) are different so scattered graph isn't appropriate.

A4_P1 We chose to use a clustered bar graph for a couple of reasons. First, using a bar graph is a great method of visually seeing comparative data. Also, it made sense to use a bar graph because we had more than one independent variable (cold temp 21 degrees C, room temp 24 degrees C, and warm temp 27 degrees C) For this reason we chose to use a bar graph as opposed to other different types of graphs. We also chose the bar graph because of the amount of data we had

Appendix 3F: Figures 3A-F: Word Clouds



BAR 2013 ADV

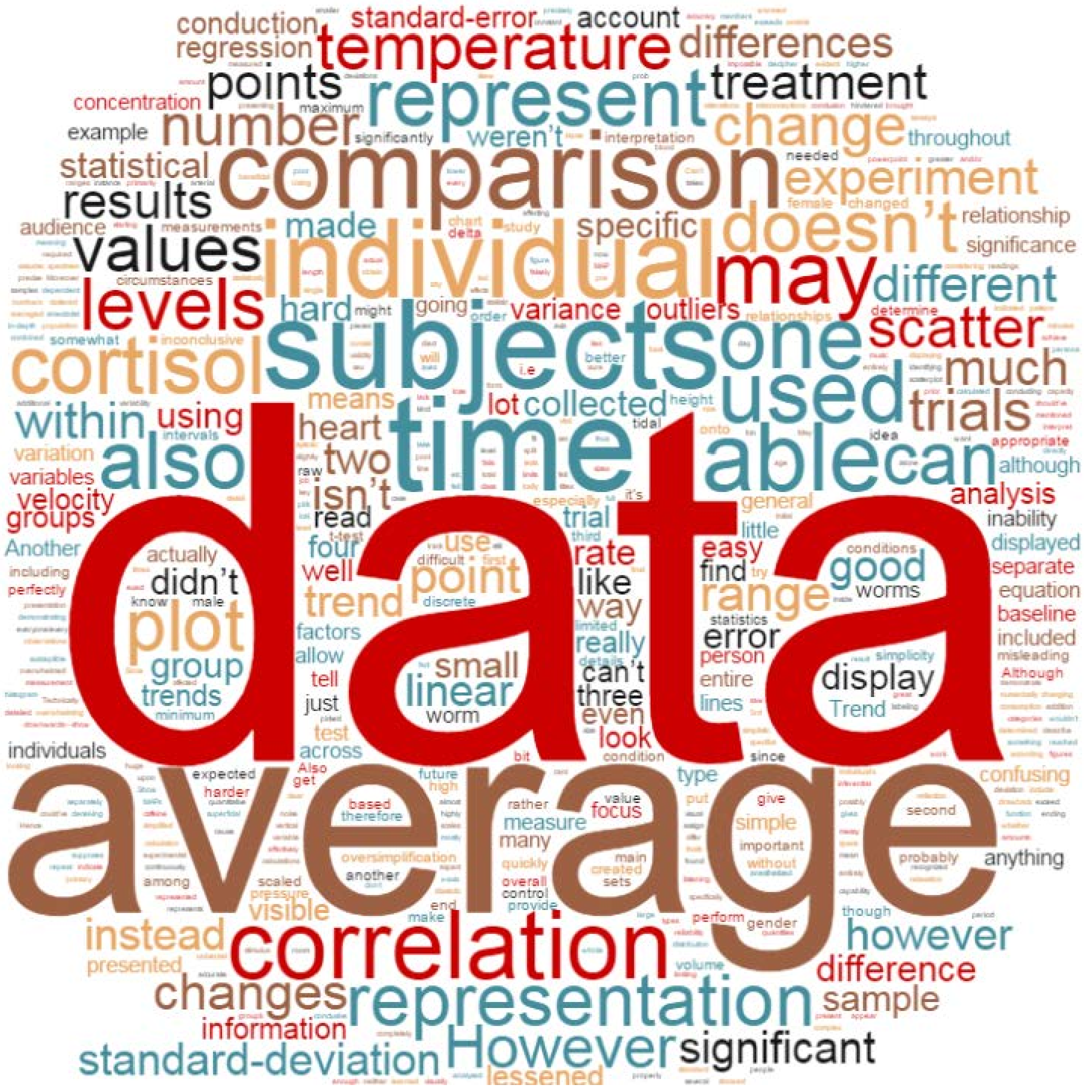
Appendix 3F continued

BAR 2014 ADV

Appendix 3F continued

BAR 2013 DISADV

Appendix 3F continued



BAR 2014 DISADV

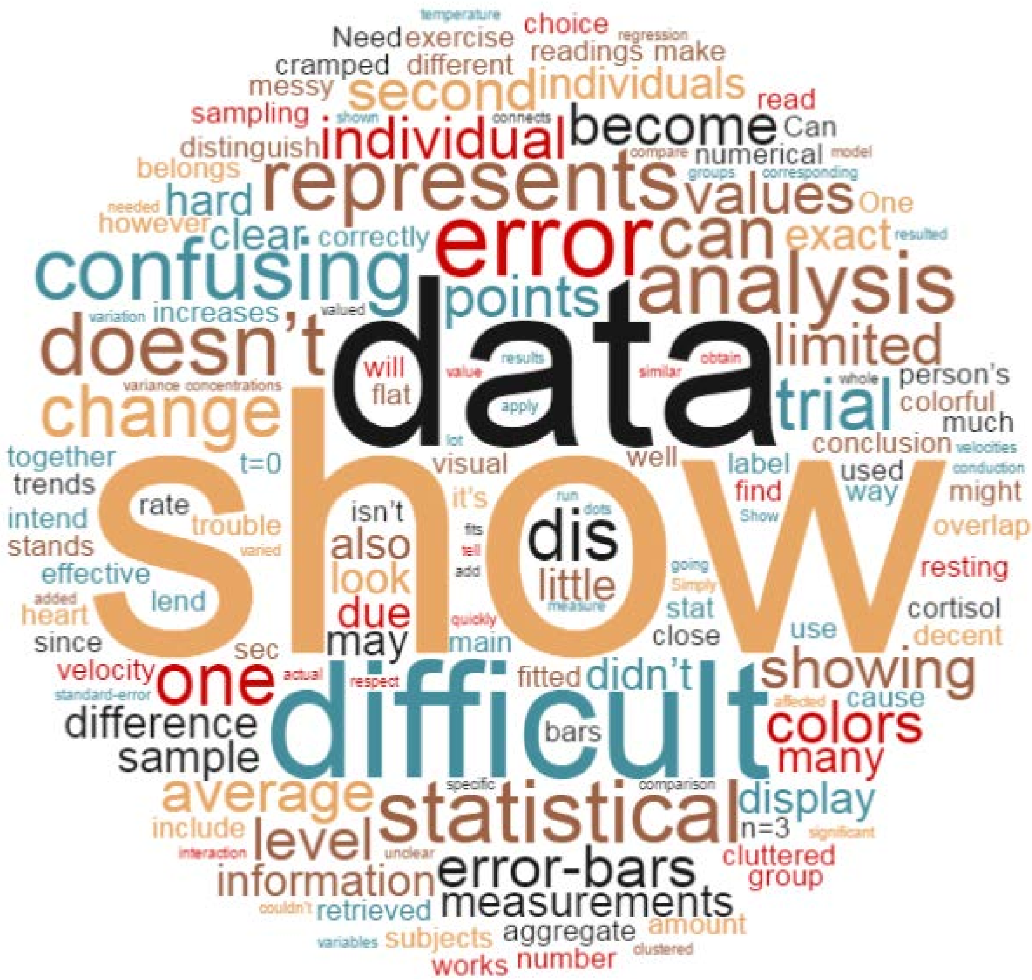
Appendix 3F continued

LINE 2013 ADV

Appendix 3F continued

LINE 2014 ADV

Appendix 3F continued



LINE 2013 DISADV

Appendix 3F continued

LINE 2014 DISADV

Appendix 3F continued

SCATTER 2013 ADV

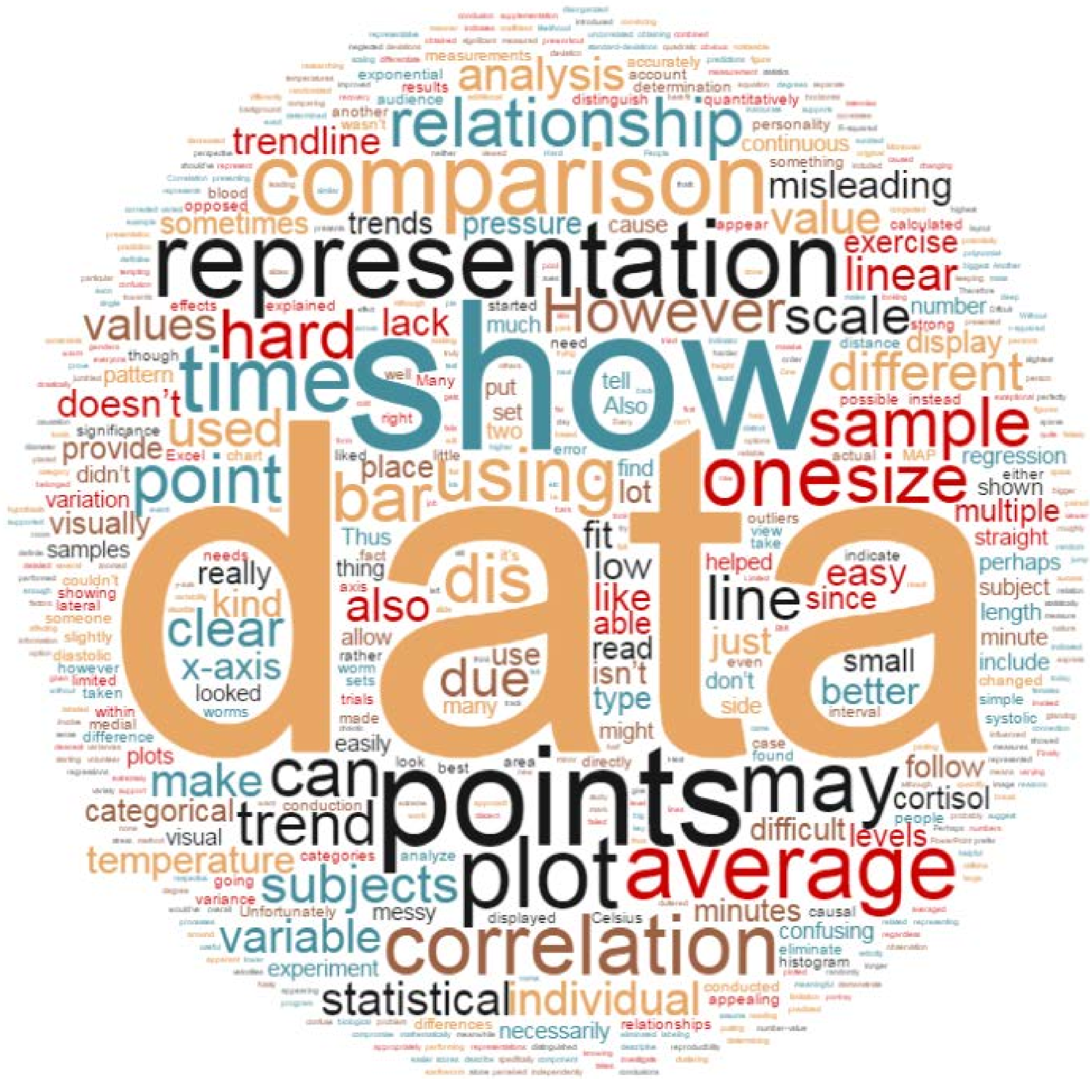
Appendix 3F continued

SCATTER 2014 ADV

Appendix 3F continued

SCATTER 2013 DISADV

Appendix 3F continued



SCATTER 2014 DISADV

Appendix 3F continued

BOX & WHISKER 2013 ADV

Appendix 3F continued

BOX & WHISKER 2014 ADV

Appendix 3F continued

BOX & WHISKER 2013 DISADV

Appendix 3F continued

BOX & WHISKER 2014 DISADV

Appendix 3F continued

DOT2013 ADV

Appendix 3F continued

DOT2014 ADV

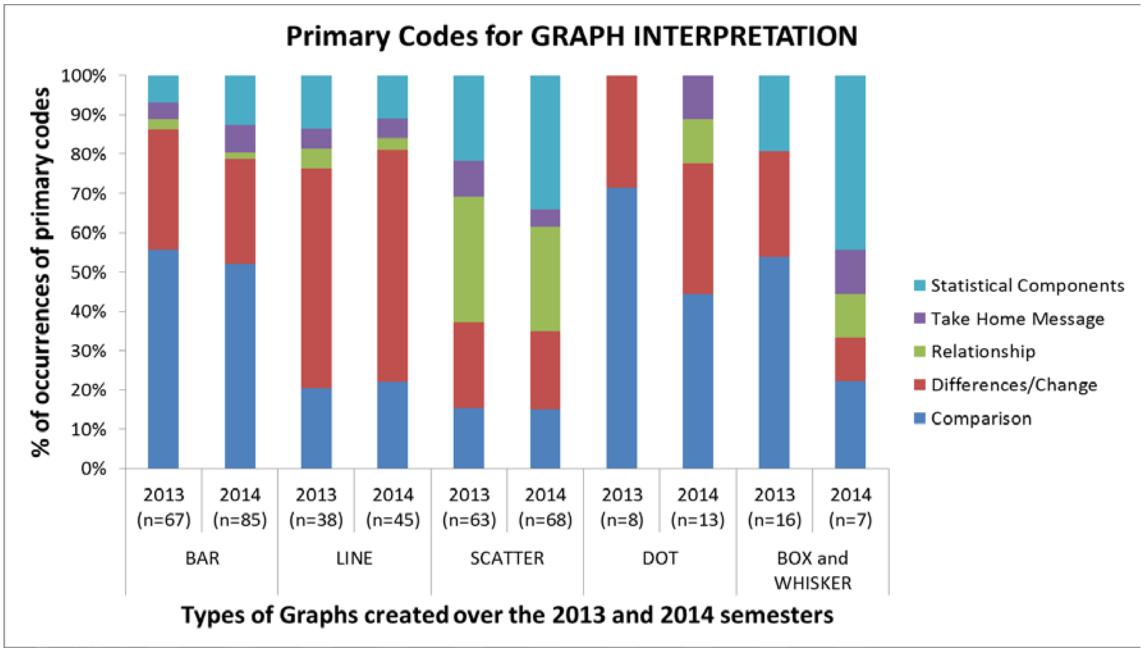
Appendix 3F continued

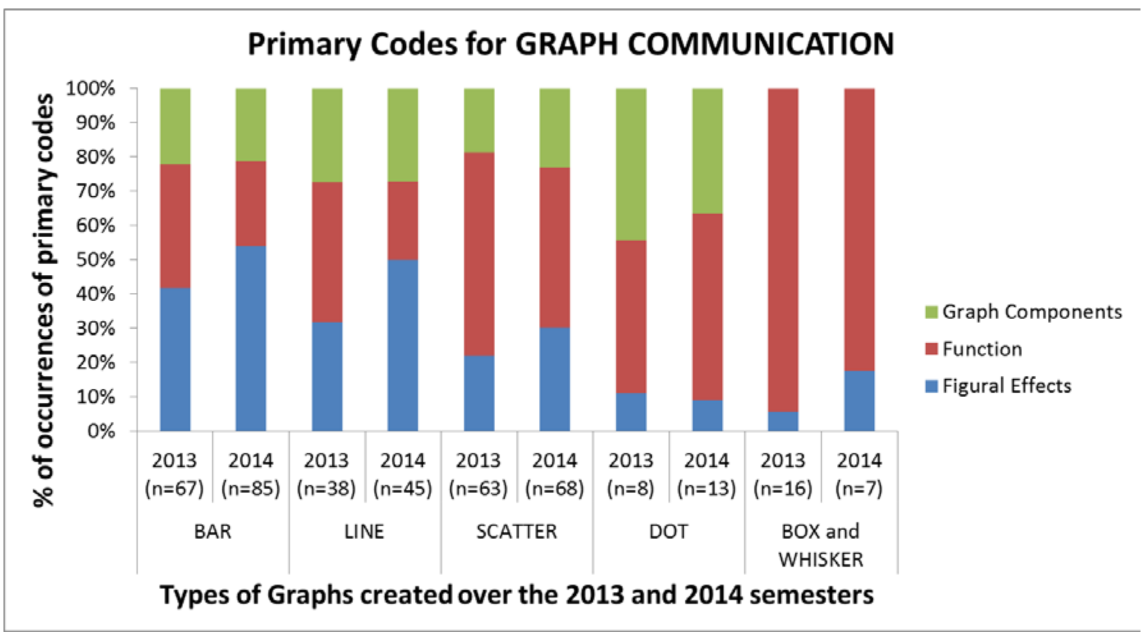
DOT2013 DISADV

Appendix 3F continued

DOT2014 DISADV

Appendix 3G: Primary Codes for Graph Interpretation and Communication for Undergraduate Students' Written Responses to Graph Choice over the Spring 2013 and 2014 Semesters





Appendix 3H: Exempt IRB consent form

PURDUE
UNIVERSITY

HUMAN RESEARCH PROTECTION PROGRAM
INSTITUTIONAL REVIEW BOARDS

To:	STEPHANIE GARDNER LILY
From:	JEANNIE DICLEMENTI, Chair Social Science IRB
Date:	08/23/2012
Committee Action:	**Exemption Granted**
IRB Action Date:	08/15/2012
IRB Protocol #:	1208012562
Study Title:	Evaluating student critical thinking through the practice of scientific inquiry

The Institutional Review Board (IRB) has reviewed the above-referenced study application and has determined that it meets the criteria for exemption under 45 CFR 46.101(b)(4) .

If you wish to make changes to this study, please refer to our guidance "**Minor Changes Not Requiring Review**" located on our website at http://www.irb.purdue.edu/policies.php. For changes requiring IRB review, please submit an **Amendment to Approved Study** form or **Personnel Amendment to Study** form, whichever is applicable, located on the forms page of our website www.irb.purdue.edu/forms.php. Please contact our office if you have any questions.

Below is a list of best practices that we request you use when conducting your research. The list contains both general items as well as those specific to the different exemption categories.

General

- To recruit from Purdue University classrooms, the instructor and all others associated with conduct of the course (e.g., teaching assistants) must not be present during announcement of the research opportunity or any recruitment activity. This may be accomplished by announcing, in advance, that class will either start later than usual or end earlier than usual so this activity may occur. It should be emphasized that attendance at the announcement and recruitment are voluntary and the student's attendance and enrollment decision will not be shared with those administering the course.
- If students earn extra credit towards their course grade through participation in a research project conducted by someone other than the course instructor(s), such as in the example above, the students participation should only be shared with the course instructor(s) at the end of the semester. Additionally, instructors who allow extra credit to be earned through participation in research must also provide an opportunity for students to earn comparable extra credit through a non-research activity requiring an amount of time and effort comparable to the research option.
- When conducting human subjects research at a non-Purdue college/university, investigators are urged to contact that institution's IRB to determine requirements for conducting research at that institution.
- When human subjects research will be conducted in schools or places of business, investigators must obtain written permission from an appropriate authority within the organization. If the written permission was not submitted with the study application at the time of IRB review (e.g., the school would not issue the letter without

Appendix 3H continued

proof of IRB approval, etc.), the investigator must submit the written permission to the IRB prior to engaging in the research activities (e.g., recruitment, study procedures, etc.). This is an institutional requirement.

Category 1

- When human subjects research will be conducted in schools or places of business, investigators must obtain written permission from an appropriate authority within the organization. If the written permission was not submitted with the study application at the time of IRB review (e.g., the school would not issue the letter without proof of IRB approval, etc.), the investigator must submit the written permission to the IRB prior to engaging in the research activities (e.g., recruitment, study procedures, etc.). This is an institutional requirement.

Categories 2 and 3

- Surveys and questionnaires should indicate
  - only participants 18 years of age and over are eligible to participate in the research; and
  - that participation is voluntary; and
  - that any questions may be skipped; and
  - include the investigator's name and contact information.
- Investigators should explain to participants the amount of time required to participate. Additionally, they should explain to participants how confidentiality will be maintained or if it will not be maintained.
- When conducting focus group research, investigators cannot guarantee that all participants in the focus group will maintain the confidentiality of other group participants. The investigator should make participants aware of this potential for breach of confidentiality.
- When human subjects research will be conducted in schools or places of business, investigators must obtain written permission from an appropriate authority within the organization. If the written permission was not submitted with the study application at the time of IRB review (e.g., the school would not issue the letter without proof of IRB approval, etc.), the investigator must submit the written permission to the IRB prior to engaging in the research activities (e.g., recruitment, study procedures, etc.). This is an institutional requirement.

Category 6

- Surveys and data collection instruments should note that participation is voluntary.
- Surveys and data collection instruments should note that participants may skip any questions.
- When taste testing foods which are highly allergenic (e.g., peanuts, milk, etc.) investigators should disclose the possibility of a reaction to potential subjects.

Ernest C. Young Hall, 10th Floor - 155 S. Grant St. - West Lafayette, IN 47907-2114 - (765) 494-5942 - Fax: (765) 494-9911

CHAPTER 4. DEVELOPMENT OF A FRAMEWORK FOR GRAPH CHOICE AND CONSTRUCTION

4.1 Abstract

Graphs are visual representations of data and play a vital role in science communication, yet past studies have shown that students in K-16 struggle with graph choice and construction. To advance graphing skills, we developed three learning and instructional materials: Guide to data displays, step-by-step guide to data communication, and the graph rubric. Here, we highlight the purpose, development, validation, and usage of these materials, which increase students' knowledge with common graphs and provide a guiding framework for data presentation.

4.2 Introduction

Collecting, understanding, and interpreting data are key skills that all students should master (Brewer & Smith 2011; Gormally, Brickman & Lutz 2012; College Board 2016; Next Generation Science Standards 2016). Research on graph interpretation and basic construction is extensive and student difficulties, primarily in K-12 type settings, have been well-documented (e.g. Graph choice (Leonard & Patterson 2004, Tairab & Al-Naqbi 2004, McFarland 2010), axes labels (Leonard & Patterson 2004, McFarland 2010), variables (Tairab & Al-Naqbi 2004), scaling axes (Ainley 2000, Brasell & Rowe 1993, McFarland 2010, Padilla, McKenzie & Shaw 1986)).

Although many instructional books exist on graphing (Bertin 1983, Tufte 1983, Kosslyn 1994, Few 2004) they do not focus on the complex reasoning behind graph choice and construction. It is insufficient to choose an appropriate graph for data (e.g. bar graph for categorical data), without evaluating the advantages and disadvantages of using a particular graph.

Metarepresentational competence (MRC) refers to the knowledge required for successful construction and reasoning with external representations, which includes graphs (diSessa 2004). MRC has four components which reveal students' ability and inability with graph choice, construction, and critique (diSessa & Sherin 2000). Specifically, these areas are invention, critique, functioning, and learning or reflection (Table 4.1; diSessa & Sherin 2000). The first area, invention, reveals students' underlying skills and abilities that allow them to conceive novel graphical representations from data (diSessa & Sherin 2000). The second area, critique, exposes students' critical knowledge that is essential for assessing various types of graphs, their strengths and weaknesses (diSessa & Sherin 2000). The third area, functioning, unearths students' reasoning for understanding the purpose of different types of graphs, and the usage being dependent on the type of data present (diSessa & Sherin 2000). The final area, learning or reflection, reveals students' awareness of their own understanding of graphs (diSessa & Sherin 2000).

Table 4.1: The table lists, defines, and connects the steps in the MRC to our work.

Steps in the MRC	**Definitions**	**Connection to graphing resources**
Invention	The underlying skills and abilities that allow students to conceive novel representations (diSessa & Sherin 2000).	Competency with graph choice, construction, and knowledge of variables is vital for conjuring new graphical representations (Leonard & Patterson 2004; McFarland 2010, Bray-Speth et al. 2010; Webber et al. 2014)
Critique	Critical knowledge that is essential for assessing the quality of representations (diSessa & Sherin 2000).	Assessing the strengths and weaknesses of various graphs exposes students' critical knowledge (McFarland 2010; Angra & Gardner 2014)
Functioning	Providing reasoning for understanding the purpose of different representations, their usage, and limitations (diSessa & Sherin 2000).	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; Angra & Gardner 2014; Webber et al. 2014)
Learning/Reflection	Strategies for fostering understanding of representations (diSessa & Sherin 2000).	Reflection, reveals students' awareness of their own understanding of graphs and gaps in their knowledge (Tanner 2012)

4.3 Overview of Graphing Literature

In addition to students, graphing difficulties have been documented in instructors (Bowen & Roth 2005), professionals (Rougier et al. 2014; Weissgerber et al. 2015), and medical doctors (Cooper, Schriger & Tashman 2001; Cooper, Schriger & Cooper 2001; Schriger & Close 2002; Schriger et al. 2006), with an interest to remediate graphing difficulties (Drummong & Tom 2011, Duke et al 2015, Saxon 2015). Although Rougier et al. (2014) suggest 10 simple rules to help with graph communication and Weissgerber et al., (2015) stress the importance of graph choice, a focus on developing the invention, critique, functioning, and learning process as suggested by MRC is lacking. Furthermore, few reports on teaching interventions and assessments of student learning around graph choice and construction exist.

While graphing is not unique to biology, it is important to consider the learning and practice of that skill within the disciplinary context to best understand and remediate biology student difficulties. Two studies at the undergraduate level in biology provide some useful insights into student difficulties and describe learning experiences in both laboratory (McFarland 2010) and lecture (Bray-Speth et al. 2010) biology classroom settings. In both studies, graphing of data was an explicit learning outcome for undergraduate biology students, but the two studies varied in numerous, and potentially important ways: classroom context, duration and focus of the intervention, the degree of scaffolding provided to students when graphing, whether students were encouraged to apply reflective and analytical skills consistent with promoting MRC, and the assessments used to evaluate student learning. McFarland (2010) designed a lab class intervention devoted to graph choice and construction which included instruction on graphs, their usage, and student engagement in the reflection on appropriateness of graphs. Throughout the semester the students were required to respond to two self-assessment prompts about their graph choice and quality, engaging an important element of MRC. There was no explicit assessment of student graphs or reasoning reported in this study. However, it is noted that faculty think the quality of student graphing improved and when students responded to a course evaluation question about their learning about graphs in the class and they reported, on average, that they learned 'a lot'. In contrast to McFarland (2010), Bray-Speth et al., (2010) administered a pre/post-test to undergraduate students in a lecture class where quantitative skills, including data and graphing, were stressed throughout the semester. Although student gains were significant in graph mechanics (e.g., title and labeled axes), some students demonstrated difficulties when tasked to choose a graph in a free response question (See Figure 4.2 in Bray-Speth et al., 2010).

The diversity in these two studies makes it difficult to deduce best practices and instructional tools to promote development of graphing proficiency. Furthermore, it is important to provide students with repeated opportunities to increase competency (McFarland 2010; Roth & Bowen 2001) and practice critical reflection in graphing choices (diSessa & Sherin 2000; diSessa 2004). There are graphing resources available to assist with graph choice and construction (Graphing Tutorial; Interactive Statistics Map; Paniello et al 2011; Puhan et al. 2006; Webber et al. 2014), however a limitation of these guides is the inefficient guidance to be reflective.

To fill the gap in graphing literature, we designed materials that are easy to implement in K-16 classrooms, are designed increase students' knowledge about graphs, and provide a systematic framework for data presentation. Here we highlight the purpose, development, and usage of three materials: (1) guide to data displays, (2) step-by-step guide to data communication, and (3) graph rubric. These materials are designed to facilitate the development of graphing competency and knowledge related to MRC.

4.4 Description and Development of Graph Materials

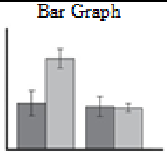
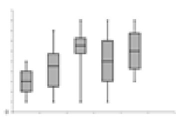
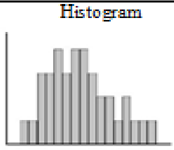
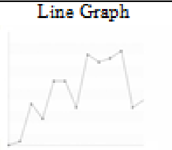
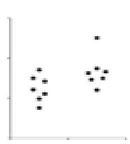
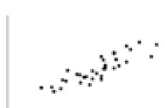
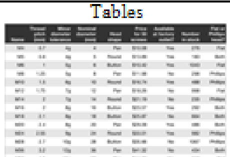
4.4.1 Guide to Data Displays

With the advent of technology and graphing software, graph makers have access to a growing number of graphic representations that they can use to construct graphs (Rougier et al., 2014). However, students and professionals default to a small subset of these graph choices, particularly bar and line graphs (Saxon, 2015; Weissgerber et al., 2015; Nuzzo, 2016). Therefore, the purpose of this resource is to increase students' MRC by exposing them to different types of graphs, their usage, advantages, and disadvantages. Undergraduate biology students are often familiar with three common types of graphs: bar, line, and scatter, and can confidently articulate the advantages of these graphs (Angra & Gardner 2014). However, when asked to articulate the disadvantages, students either display uncertainty on elements that comprise disadvantages, or they naively state no disadvantages (Angra & Gardner 2014). Critical graph evaluation is vital for refining critical thinking and argumentative skills, and is an important part of MRC (diSessa 2004; Kuhn & Udell 2003; McFarland 2010).

The guide to data displays (Table 4.2) was inspired by an earlier draft used within the context of an upper-division physiology course, (See Appendix 3B in chapter 3), whose purpose was to expose students to different types of graphical representations. The earlier version provided real-life examples of graphs from textbooks, science journals, and stated the purpose of each type of graph. What it did not do was list the advantages and disadvantages of each. In chapter 3, students were asked to reflect on their graphs constructed in lab, particularly the advantages and disadvantages. After conducting a recursive analysis (see chapter 3), we found it necessary to inform students of the advantages and especially the disadvantages, in order to increase students' MRC for choosing and critiquing graphical representations.

The guide to data displays (Table 4.2) displays six common types of graphs (bar, box and whisker, histogram, line, dot, scatter) and tables used in the biological sciences, along with relevant citations that describe the usage, advantages and disadvantages of each graph. The decision to select these graphs was obtained from: a) surveys from professors (Angra & Gardner 2014) b) think-aloud interviews with expert professors and novice students (Angra & Gardner 2014); c) data collected from an upper-level laboratory classroom at a large midwestern university (Angra & Gardner 2014); d) seminal literature, which included relevant books, undergraduate biology textbooks, articles from various fields of biology and education literature. All six graphs are bound by the Cartesian coordinate system which allows graphs to portray relationships between variables. Although students are exposed to pie charts on a daily basis (phone apps, magazines, television, etc.) (Angra & Gardner 2014), we found minimal presence of pie charts in biology textbooks and primary literature. This could be because pie charts are used to display frequency or percentage data, which is useful for understanding emerging patterns in data (Kosslyn 1994; Schriger & Cooper 2001), but they cannot communicate relationship between experimental variables.

Table 4.2: Summary of common graphs, their usage, advantages, and disadvantages. (Figure for the dot plot was taken directly from Weissgerber et. al, 2015.)

Data Display Type	Usage	Advantages	Disadvantages
Bar Graph 	To compare categorical data, percentages, or summary statistics from multiple groups. [66] Each bar represents a category; shape can be changed by moving the categories around.[39]	Useful for understanding distributions from large datasets. [66]. Stacked bars or shading of bars can be used to distinguish the different levels within the data.[66]	Obscures the distribution of data[66, 76], number of data points, and their values.[25, 26]
Box and Whisker Plot 	To show distribution of data from one or multiple groups[26, 66].	Shows and compares distributions of large datasets.[26, 66]	Should not be used for small datasets.[66] Does not show individual data(except for outliers).[66]
Histogram 	To show a distribution of data with the independent variable as continuous.[39] Uses numerical data instead of categorical data[39].	Shows the shape of the distribution of data with a continuous variable.[39]	Must choose the bin size wisely to avoid influencing the shape being too compressed or too dispersed.[39]
Line Graph 	To show how a single variable or multiple variables changes over time or to show how a variable deviates from a set baseline[30] X axis portrays categories while the Y axis portrays quantitative values.[30]	Shows direct relationships and may be used to predict relationships between continuous variables.[51]	Not appropriate for representing ranked, part-to-whole, or correlational data.[30]
Dot Plot 	To show distribution of small data sets from multiple groups .[25, 76] The independent variable is categorical and the dependent variable is continuous.	Shows all data from multiple categories and the distribution within each category.[25, 76]	Not appropriate for representing a large data set because the plot will become cluttered and it will be difficult to see the individual points.[25]
Scatterplot 	To show individual data points from bivariate data.[66]	Shows the relationship between variables. [26, 66] Shows trends in the data and any noticeable outliers.	It may be difficult to extract individual data points if they fall on the same or nearby coordinates.[26, 66]
Tables 	To display simple relationships between numerical values and categorical groups, so that individual values can be easily extracted from the rows and columns[30]. Often used for small data sets. [28, 72]	Since values in a table are encoded as text, it is easy to extract individual values. [28, 30] Numbers in a table can be displayed with decimal precision.[30, 72] A table can also communicate multiple sets of data with different units. [30]	Tables may make it difficult to interpret the take home message if not organized properly.[28]

4.4.2 Step-by-Step Guide

Previous findings from our expert-novice study (chapter 2) revealed that undergraduate students did not think of data in the same way as professors when constructing their graphs. Unlike the professors, some of the undergraduate students did not actively plan or talk about the purpose of the graph prior to construction or reflect on their graph post-construction. Therefore, the purpose of this guide (Table 4.3) is to provide students with a framework for data presentation, which encompasses all four components of the MRC (Table 4.1) and allows students to practice expert-level reasoning.

Under an approved IRB protocol (#1210012775, see Appendix 2F), undergraduate students (n=6) and professors (n=7) were recruited from within the biological sciences for semi-structured, think-aloud interviews (See chapter 2). Participants were given a simple data set and were asked to plot the data graphically. Graphs were constructed by hand using a Livescribe pen and notebook paper instead of digitally on the computer. The Livescribe pen synchronizes written notes with recorded audio and has an embedded infrared camera that detects penstrokes when used with the Livescribe dot paper (Livescribe 2006). The usage of this pen allowed us to understand the step-wise reasoning process used during graph construction and enabled the development of the step-by-step guide. A summary of the thought process for undergraduate students and professors is illustrated in Figure 4.1. We reached this summary of the thought processes (Figure 4.1) by listening to and watching the individual audio and video files from the think-aloud interviews and ordering the steps in which they occurred as the participant proceeded with their thinking in terms of data, graphing, and reflection, summing across each participant group, averaging the step and ordering it. Not all professors and not all students followed all of the steps shown in Figure 1, but shown are the average step combinations.

Figure 4.1 shows that professors took 12 steps to construct their graph, while undergraduate students took only 8 steps. However, given the purpose of this graphing tool, to scaffold student learning to that of the experts, a top-down approach was taken to fill in the gaps. Ordered data from the think-aloud interviews and four additional steps informed the final order of 16 steps. These steps were further categorized into three phases: Planning, execution, and reflection (Table 4.3). Terminology for the phases are adapted from Koedinger and Anderson's (1990) Diagram Configuration model used for solving geometry proofs and from Polya's problem solving cycle in mathematics (Polya 1945). Polya's first principle challenges the learner to understand the problem, which is similar to our planning phase, consisting of formulating a purpose for the graph and organizing the data. Polya's second principle instructs the learner to devise a plan. This principle is also embedded in our planning phase, consisting of classifying variables, deciding on data manipulations, and finalizing the graph choice. Polya's third principle instructs the learner to carry out the plan and parallels our execution phase, and diSessa's invention step in the MRC where the learner actively constructs the graph with appropriate mechanics. Polya's last principle instructs the learner to look back through the problem, check the result, and reflect on the problem solving approach. This step closely resembles our reflection phase and diSessa's critique, functioning, and reflection steps in the MRC and consists of critical reflection on graph alignment, graph choice, and data presentation. The similarity in our work, Polya's work in mathematics, and diSessa's work on MRC provided the impetus to develop common steps which engages and challenges the learner to develop systematic independent and intellectual skills to solve problems. Differences lie in the nature of the task. Polya's work was developed in the context of mathematics and the four principles apply broadly to all problems students may encounter in mathematics courses. DiSessa's work is rooted in cognitive psychology to understand how students construct and interpret scientific representations. The materials we developed are focused even further on only graphical representations.

Our analyses comparing the step-wise process that occurred during graph construction between professors and undergraduate students reveal two things (Figure 4.1). First, professors take four more steps during graph construction compared to undergraduate students. Specifically, after professors read the prompt, they spent time planning and articulating their research question and variables before constructing their axes (See Figure 2.1 in chapter 2). Unlike the undergraduate students who decided on a graph type early on in the interview, this step for the professors occurred much later on, after they had labelled their axes. Sometimes, professors did not articulate their graph choice, but automatically chose a graph type and proceeded with the construction, This could be because when an expert is solving a problem, they are executing several subprocesses that vary from fluent to automatic (Bransford , Brown, and Cocking, 2000). By focusing their conscious attention on the variables and data presented in the task, we think that expert professors let the data guide their choice of graph (Bransford, Brown, and Cocking, 2000). The second revelation is that when professors are nearing the end of their graph construction, they take time to reflect on their creation by thinking of other ways to graph data, the disadvantages of their creation, and they provide an interpretation of their graph by including a figure legend. The automatic reflection present during the construction task could be due to professors' extensive experiences with experiments, data analysis and graph construction in their own research and teaching others (Chi, Feltovich & Glaser 1981; Roth & Bowen 2001).

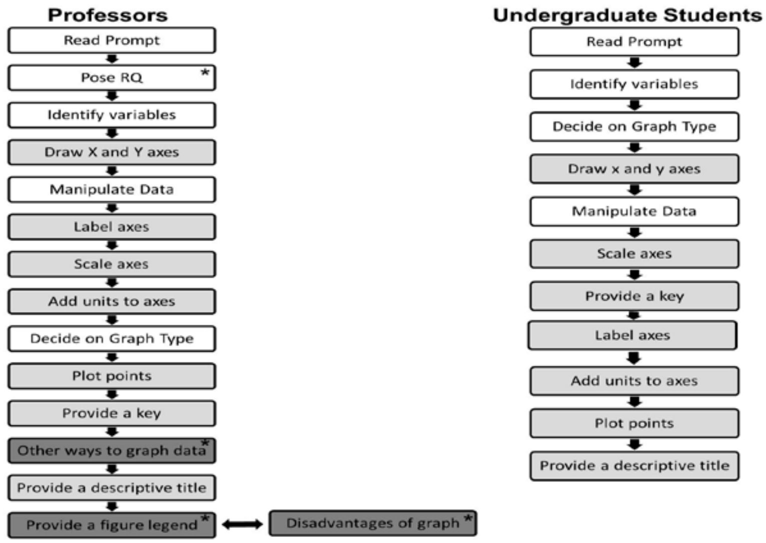


Figure 4.1: Comparing thought processes between professors and undergraduate students: Data from think aloud interviews reveals the underlying thought processes used by expert professors (n=7) and undergraduate students (n=6) when translating a table of raw values into a graph. There were 16 steps in total with 12 being taken by professors and 8 by the undergraduate students. White boxes illustrate the planning phase, which consists of organizing data, deciding on the purpose of the graph, and deciding on a graph type. Light gray boxes represent the execution phase, which consists of steps needed for graph construction. Dark gray boxes designate the reflection and explanation phase which comprises of critical reflection of graph choice, and take home message. The asterisks indicate the extra steps that professors took during graph construction that were not taken by undergraduate students during graph construction in our think-aloud interviews. The bidirectional arrow indicates the steps taken by professors as either their last or second to last step in graph construction.

Table 4.3: Step-by-step guide to data communication.

Phases 1-3	Elements	Notes for your Experiment
1. Planning- In this phase, you must organize your data and decide on the message you want to communicate in your graph. It helps to first conceptualize the whole task before executing it.	Step 1- Revisit your research question and hypothesis and ask yourself, what is it that you want the graph to show?	
	Step 2- Identify your independent and dependent variables.	
	Step 3- Classify your variables as either categorical or continuous.	
	Step 4- Decide whether or not you need to manipulate your data.	
	Step 5-Decide on a graph type that will best represent your data.	
2. Execution- In this phase, you will actively construct a graph.	Step 6-Label the axes with your variables.	
	Step 7-Add units to the axes, if necessary.	
	Step 8- Adjust the scale of axes into appropriate increments for the data.	
	Step 9-Include a key, if appropriate.	
	Step 10-If you are displaying the graph in a report, include a figure legend.	
	Step 11-Include a descriptive title.	
3. Reflection and Explanation Phase- In this phase, you will critically reflect on your graph choice, interpret your graph, and explain your answers to questions posed in steps 13-16.	Step 12-Check the alignment of your representation with your research question and hypothesis.	
	Step 13-What are the advantages of the representation?	
	Step 14- What are the disadvantages of the representation?	
	Step 15- What is the take-home message of the representation?	
	Step 16-What are some other ways that you could have represented your data?	

4.4.2 The Graph Rubric

Knowing how to construct appropriate graphs that fit the data and convey the take home message is an essential skill, but difficult to acquire without proper guidance and feedback. Findings from our previous studies with undergraduate students (chapters 2 and 3) and those in the literature at the K-12 and undergraduate levels have documented difficulties with graph mechanics, choice, and alignment with the research question and hypothesis (Padilla et al., 1986; Brasell & Rowe, 1993; Ainley, 1995; Mevarech & Kramarsky, 1997; Bakker, 2004; Leonard & Patterson, 2004; Tairab & Al Naqbi, 2010; Meletiou & Lee, 2010; Speth et al., 2010; Hattikudur et al., 2012; Angra & Gardner, 2016). One of the reasons that these difficulties remain with the undergraduate students is because there isn't explicit instruction on data representation at the higher education levels. Rubrics are valuable tools in the classroom because they make learning goals explicit for both the instructors and students (Allen & Tanner, 2006; Wolf & Stevens, 2007). This article motivated us to develop a valid and reliable analytic graph rubric for our research and for classroom use. Particularly, we wanted to design a tool to help instructors provide quick and consistent feedback on students' graphs and for students to use when constructing and critiquing graphs. The graph rubric was modified drastically from an earlier draft used in a naturalistic classroom study (See Appendix 3C in chapter 3). The earlier version of the graph rubric consisted of 4 categories, whereas the current rubric consists of twelve categories with detailed explanations for what accounts as an "excellent, needs improvement, or unsatisfactory". Outlined below are the steps that occurred for development and validation of this rubric.

4.4.3 Development of the Graph Rubric

For the creation of the graph rubric, we consulted the graphing literature (see Table 4.4 for citations), student-generated graphs and reflections from a classroom study (Angra & Gardner 2014), and graphs constructed by students and professors in a clinical interview (Angra & Gardner 2014). A well-designed graph consists of many different components, each vital for graph interpretation (Angra & Gardner 2016). In order to organize the different components in a comprehensible manner, a rubric with three levels of achievement (excellent, needs improvement, and unsatisfactory) was created with three main categories: (1) graph mechanics, (2) communication, and (3) graph choice. Graph mechanics consists of elements that must be present in graphs: descriptive title, axes labels, units, scale, and key. Communication is divided into the ease of understanding given the aesthetics of the graph and the basic take home message. Graph choice consists of the appropriateness of the graph type, data type plotted, and the alignment of the graph with the research question and hypothesis. Descriptions and citations of the different categories in the rubric can be found in Table 4.4.

Table 4.4: The graph rubric with citations.

	Rubric Categories	Citations
Graph Mechanics	**Descriptive title** • P/A-Should be: a) in the form of a statement, b) mention the subject, c) appropriate variables, and d) include relevant details about the experiment that help understand the take home message. • NI- If the title is missing any one of the four points mentioned above.	Kosslyn, 1994 College Board, 2012
	Label for the X axis (e.g. time) • P/A- Should be appropriate and descriptive for the experiment. For graphs with categorical independent variables, there needs to be a label under each set of data and a larger label under all data plotted. • NI- If the label is missing any one of the points mentioned above.	Schriger & Cooper, 2002 College Board, 2012 Humphrey et al., 2013 Cooper et al., 2001 Kosslyn, 1994 DiFazio, 1990 Elliot et al., 2006 Leonard & Patterson, 2004 Patterson & Leonard, 2007 Ainley, 2000 Brasell & Rowe, 1993 Bruno & Espinel, 2009 Leinhardt et al., 1990 Berg & Phillips, 1994
	Label for the Y axis (e.g. heart rate) • P/A- Should be appropriate and descriptive for the experiment. If the data is manipulated (average, change, percentage, etc.), then it should be indicated on the y axis. • NI- If the label is missing any one of the points mentioned above.	Schriger & Cooper, 2002 College Board, 2012 Humphrey et al., 2013 Cooper et al., 2001 Kosslyn, 1994 DiFazio, 1990 Elliot et al., 2006 Leonard & Patterson, 2004 Patterson & Leonard, 2007 Ainley, 2000 Brasell & Rowe, 1993 Bruno & Espinel, 2009 Leinhardt et al., 1990 Berg & Phillips, 1994
	Units for the X axis (e.g. seconds) • P/A- Should be appropriate and descriptive for the data displayed. • NI- If the units are not appropriate or descriptive.	Elliot et al., 2006 Leinhardt et al., 1990
	Units for the Y axis (e.g. average beats per minute) • P/A- Should be appropriate and descriptive for the data displayed. • NI- If the units are not appropriate or descriptive.	Elliot et al., 2006 Leinhardt et al., 1990
	Scale (appropriate intervals and range for data) • P/A- Should be appropriate for the data displayed such that the increments are clear and without clutter and includes appropriate significant figures. If the scale is discontinuous or doesn't start at the origin, it should be indicated by a break in the axis. • NI- If the scale is not appropriate for the data such that it is cluttered, does not include appropriate significant figures, and/or if the scale does not indicate axis break.	Schriger & Cooper, 2002 College Board, 2012 Humphrey et al., 2013 Franzblau & Chung, 2012 Cooper et al., 2001 Kosslyn, 1994 DiFazio, 1990 Patterson & Leonard, 2007 Ainley, 2000 Brasell & Rowe, 1993 Bruno & Espinel, 2009 Leinhardt et al., 1990 Berg & Phillips, 1994
	Key (defines different data sets that are plotted) • P/A- Should be appropriate and descriptive for the data displayed. It should include: a) descriptions of different colors (if applicable), b) the sample size and c) the number of trials. • NI- If the key is not descriptive and does not indicate the sample size.	Schriger & Cooper, 2002 Kosslyn, 1994 Quinn & Keough, 2002

Table 4.4 continued

		Citations
Communication	**Ease of Understanding-Aesthetics** •E- If the graph is aesthetically pleasing, meaning that: a) the data plotted takes up sufficient room in the Cartesian plane, b) makes use of legible size font, c) the x and y axis lines are clear and legible, d) the graph displays data in an appropriate number of bars and lines, and e) is devoid of chart junk elements such as: distracting background colors, patterns, and dark gridlines •NI- If the graph has one of the following flaws: a) the graph displays too much white space, b) the font size is too small, c) the x and y axis lines are not clear and legible, d) the graph shows too many bars or lines OR e) elements of chart junk are clouding interpretation of data. •U- If the graph has multiple flaws, which interfere with the understanding and interpretation of data.	Rougier et al., 2014 Tufte, 1983 Schriger & Cooper, 2002 Franzblau & Chung, 2012 Kosslyn, 1994 Patterson & Leonard, 2005 Quinn &Keough, 2002 Few, 2011
	Ease of Understanding-Take home message • E- If the graph has sound construction and mechanics that allow for clear sorting of trends and take home message. • NI- If data trends are difficult to observe or it is difficult to formulate a proper take home message. • U- If the graph is ineffective at communicating data trends and take home message, such that it causes confusion.	Rougier et al., 2014 Schriger & Cooper, 2002 Franzblau & Chung, 2012 Kosslyn, 1994 Patterson & Leonard, 2005 Ainley, 2000 Few, 2011
Graph Choice	**Graph Type (Bar, line, scatter, dot, box and whisker)** •E- If data displayed in a graph is appropriate for both independent and dependent experimental variables (i.e. categorical and continuous) and data. (*Referring to the data form) •NI- If data displayed in a graph is a) not suitable for either the dependent or independent experimental variables OR b) there is a better way to present data. •U- If the graph type is not suitable for both experimental variables.	Schriger & Cooper, 2002 College Board, 2012 Humphrey et al., 2013 Franzblau & Chung, 2012 Kosslyn, 1994 Leonard & Patterson, 2004 Patterson &Leonard,2005 Patterson & Leonard, 2007 Tufte, 1983 Bruno & Espinel, 2009 Quinn & Keough, 2002
	Data Displayed (Raw, Averages, Changes, Percentage) •E- If the graph indicates the type of data (ex. Raw, averages, etc.) that are plotted. There should be a clear distinction between raw data and manipulated data based on the information presented in the key (ie. sample size and number of trials) and axis label. If the graph is showing averages, then it should also be accompanied with STDEV or error bars. •NI- If the graph is missing one of points mentioned above. •U- If data type is inappropriate for the graph type	Schriger & Cooper, 2002 Humphrey et al., 2013 Kosslyn, 1994 Leonard & Patterson, 2004 Quinn & Keough, 2002
	Alignment *(at least one of the graphs presented should align with the research question and hypothesis. Other graphs can be exploratory.)* •E- If the graph is completely aligned with the research question and/or hypothesis. In other words, the independent, dependent variables, and information about the experiment are explicit. •NI- If the graph is partially aligned with the research question and/or hypothesis. In other words, the graph is missing information about either the independent, dependent, or details about the experiment. •U- If the graph is not aligned with the research question and/or hypothesis.	Schriger & Cooper, 2002 Franzblau & Chung, 2012 Kosslyn, 1994 Patterson & Leonard, 2005 Ainley, 2000 Rougier et al., 2014 Quinn & Keough, 2002

During the development processes and early rubric drafts, we solicited informal feedback from undergraduate students, graduate students, and professors from science education and the biological sciences. Incorporating feedback from instructors (graduate students and professors) at various levels of education and from various fields of expertise provided us with broad applicability. Feedback from students allowed us to make sure that the language in the rubric was clear and easy to understand.

The final version of the graph rubric was validated three different ways and by three different groups of people: 1) undergraduate students used the graph rubric to critique graphs constructed in the classroom; 2) professors used the rubric to validate graphs present in biology journals; and 3) AA and SMG used the rubric to critique graphs found in introductory biology textbooks.

4.4.4 Methods of Graph Rubric Validation

Under an approved IRB protocol #1210012775 (see Appendix 2F), emails were sent to undergraduate students who successfully completed an upper-division physiology course in which experimental design, data collection, and communication were heavily emphasized. Personal emails were also sent to professors affiliated with the biological sciences and those who previously participated in our think-aloud, clinical interviews. The final sample size for the undergraduate students was 7 and for professors was 3. Participants were remunerated with a $10 Amazon gift card.

Validation of the graph rubric was conducted remotely by undergraduate students and professors and all communication was done via email. In order to guide students and professors during their graph critique, a graph rubric guide was constructed (see

Appendix 4H). This rubric guide defined, explained, and provided examples of five graphs, each from the three levels of achievement.

The undergraduate students were asked to critique five graphs (Table 4.6) that represented the types of graphs constructed over the course of the Spring 2015 semester and accounted for common graph errors made by undergraduate students in the classroom context. Professors were asked to send three research articles from their sub-discipline in biology to AA, who then randomly selected five graphs (Tables 4.7, 4.8, and 4.9). Since graphs are the most common visual representation in research articles and are usually accompanied by other forms of visual representations (e.g. tables, images taken at a macro and micro level, images from gel electrophoresis, diagrams, etc.), professors were given the entire figure panel with figure legend as it often contains important information for graph interpretation such as sample size, key, and information on the subject and experiment (Rybarczyk, 2011).

Students and professors were instructed to consult the rubric guide first and then proceed with graph critique. Students and professors were also encouraged to comment and explain their reasoning for scoring graph rubric categories. All graphs were approved by SMG before they were sent for critique. Professors and undergraduate students were also asked to fill out a follow up post graph evaluation survey in order to get their opinion on the graph rubric and its usability. Please see Appendix 4I for undergraduate students and Appendix 4J for professor responses. To quantify the degree of agreement, inter-rater reliability (IRR) was calculated using Excel (McHugh, 2012). Since we did not train the undergraduate students or professors on the graph rubric, we also calculated Fleiss' kappa in R (R Core Team, 2013) to account for random guessing that may have occurred when scoring graphs with the rubric with the 7 undergraduate students, Aakanksha Angra and Stephanie M. Gardner, from here on referred to as AA and SMG, respectively. Fleiss' kappa was calculated in R (R Core Team, 2013) with professors and followed up with a Cohen's kappa conducted in Statistical Package for the Social Sciences (SPSS, V.22). We used suggestions by Landis and Koch (1977) to categorize our IRR agreement values. Kappa values range from -1.0 to 1.0, where a value of 1 indicates perfect agreement, a value of 0 is indicated by chance, and values below 0 indicate values less than chance (Viera & Garrett, 2005). We use the metrics put forth by Fleiss (1981) to categorize the strength of agreement for Fleiss's kappa values. Values greater than 0.75 represent "excellent" agreement beyond chance; values between 0.4-0.75 represent "fair-good" agreements, and values below 0.4 represent "poor" agreement.

We used kappa metrics for Cohen's kappa suggested by Landis and Koch (1977) to categorize the strength of agreement for the Cohen's kappa values: values that are above 0.81 are considered "almost perfect"; values that range between 0.61-0.80 indicate a "substantial" level of agreement; values between 0.41-0.60 are considered "moderate"; values between 0.21-0.40 are "fair"; values between 0.0-0.20 are "slight"; and everything below 0.0 is "poor" level of agreement. Per the recommendation of McHugh (2012), and strengths and weaknesses of each type of statistic, we report the results from the IRR, Fleiss' and Cohen's kappa tests.

This rubric was developed to improve graphing at the undergraduate level. When students are not actively constructing graphs, they are exposed to them in their classes primarily through textbooks (Angra & Gardner, 2014). With the goal of validating graphs in textbooks, we chose five introductory biology textbooks, listed in Table 4.5.

Four textbooks (Raven et al., 2008; Sadava et al., 2009; Urry et al., 2014; Singh-Cundy & Shin, 2015) were chosen based on the undergraduate curriculum for biology students at a large midwestern university and the fifth textbook (Integrated Concepts in Biology; Campbell et al., 2011) was chosen because of its carefully constructed content which directly addresses the recommendations put forth by Vision and Change to incorporate more quantitation in biology (Brewer & Smith, 2011).

In order to evaluate graphs from biology textbooks, we based our selection criteria from a paper published by Rybarczyk (2011). We used a random number generator to randomly select ten chapters from each textbook. The chapters were examined and pages that displayed graphs were scanned. Each graph was analyzed as a stand-alone figure with the graph rubric. The definition that we use for a graph is taken from Kosslyn's work (1994) as "a visual display that illustrates one or more relationships among numbers". We expanded this definition and analyzed graphs that were in a Cartesian coordinate system, framed with an x and y axis, and found in the main chapter or in the side panel chapter exercises (see Appendix 4I for examples). We excluded interactive graphs, graphs found in videos, and graphs found at the end of the chapter exercises.

The final count of the graphs analyzed is displayed in Table 4.5. Since the purpose of the Integrated Concepts in Biology textbook is to incorporate more experiments, data, and quantitation in biology, we noticed that compared to the other four textbooks analyzed, on average, there were approximately seven times more graphs present in this online textbook. For the purpose of validating our rubric and staying consistent with the number of graphs present in the other 4 traditional textbooks, we analyzed twenty percent of the graphs, chosen at random. To quantify the degree of agreement assessing the textbook graphs, we report IRR and not Cohen's kappa because both raters AA and SMG developed the rubric, had prior training providing feedback to students using a similar version of the rubric, and previously reported "strong-almost perfect" levels of agreement (McHugh, 2012). Since graphs in textbooks are not accompanied by research questions or hypotheses, we could not include the "alignment" category in our IRR.

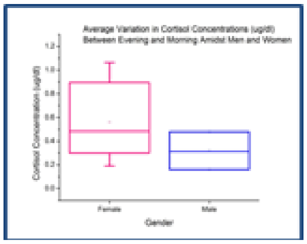
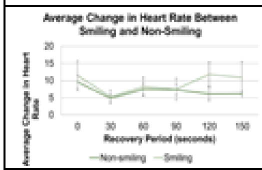
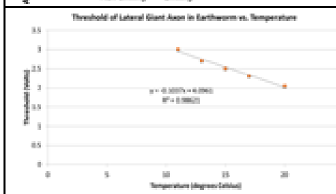
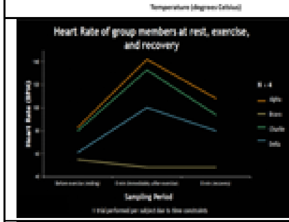
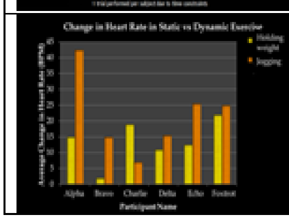
Table 4.5: Five introductory biology textbooks used for graph rubric validation

Textbook	Randomly Selected Chapters	Number of Pages Analyzed	Number of Graphs Analyzed
Singh-Cundy & Shin, 2015	2, 3, 19, 20, 22, 26, 27, 29, 30, 33	227	17
Sadava et al., 2009	5, 11, 14, 28, 38, 45, 52, 53, 56, 57	207	35
Urry et al., 2014	3, 4, 6, 11, 13, 18, 22, 29, 32, 38	211	18
Raven et al., 2008	20, 23, 25, 28, 34, 44, 45, 52, 58, 59	203	43
Campbell et al., 2011	1, 2, 5, 9, 10, 15, 22, 23, 26, 29	n/a online textbook	204 graphs present, but analyzed 40 graphs (20%)

4.4.5 Results from Graph Rubric Validation

Since one of the goals of the graph rubric is for it to be used as an assessment tool, it was imperative post-development that it be validated both with expert professors and undergraduate students to ensure utility and difficulties it was designed to resolve. Results from the first phase of validation with nine raters (seven undergraduate students, AA, and SMG) evaluating five graphs constructed in the classroom are shown in Table 4.6. IRR and Fleiss's data are reported for the graph mechanics, communication, and choice (Table 4.6). IRR for graph mechanics ranged between substantial (graphs1 and 2) to almost perfect (graphs 3-5). IRR for graph communication fell into three categories: moderate (graphs 3 and 5), substantial (graph 4) to almost perfect (graphs 1 and 2). IRR for graph choice ranged between substantial (graphs1, 2, 4, and 5) to almost perfect (graph 2). Fleiss' kappa ranged between fair-moderate for the graph mechanics category, but was more variable for graph communication, ranging from poor to substantial level of agreement, and for graph choice, ranging from poor to moderate levels of agreement.

Table 4.6: Average IRR and Fleiss' Kappa across five graphs for 9 raters: 7 undergraduate students, AA, and SMG.

Graph Evaluated using the Graph Rubric	Graph Mechanics		Graph Communication		Graph Choice	
	IRR	Fleiss' Kappa	IRR	Fleiss' Kappa	IRR	Fleiss' Kappa
	0.77	0.41	0.88	0.60	0.63	0.23
	0.77	0.50	0.83	-0.07	0.74	0.08
	0.82	0.43	0.50	0.22	0.81	0.41
	0.90	0.35	0.66	0.00	0.62	-0.01
	0.92	0.58	0.55	-0.13	0.62	-0.18

Results from the second phase of validation with three raters (professor, AA, and SMG) evaluating five graphs from research articles from within the sub-disciplinary field of the professor are shown in Tables 4.7, 4.8, and 4.9. Overall, compared to the undergraduate students, professor agreements across graph mechanic, communication, and choice showed either a “substantial” or “almost perfect” agreement. IRR values were the highest for the graph mechanics category, followed by communication and graph choice. Professor 2 had the highest occurrences of perfect IRR. Since values for Fleiss’ kappa fell into all categories suggested by Landis and Koch (1977), we performed a Cohen’s kappa between two raters: AA-SMG, professor (P)-AA, and P and SMG. Results from Cohen’s kappa are displayed in Table 4.10. The highlighted value in each cell indicates the highest agreement of the three pairs.

Table 4.7: Average IRR and Fleiss' Kappa values for 5 graphs with Professor 1, AA, and SMG.

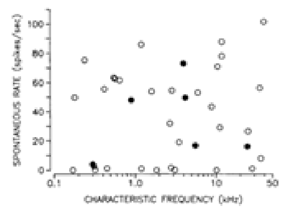
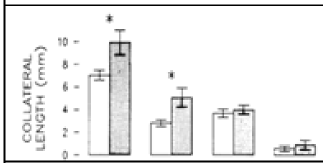
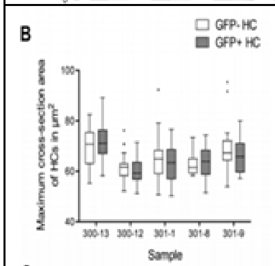
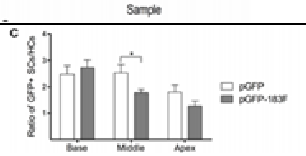
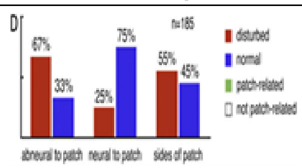
Rubric Validation with Professor 1						
Graph Evaluated using the Graph Rubric	Graph Mechanics		Graph Communication		Graph Choice	
	IRR	Fleiss' Kappa	IRR	Fleiss' Kappa	IRR	Fleiss' Kappa
	0.85	0.074	1	1	0.77	0
	0.85	0.47	0.83	0.25	0.89	0.36
	0.90	0.62	0.67	-0.33	0.67	0.1
	0.85	0.39	0.83	-0.3	0.67	0.1
	0.81	0.43	1	1	0.67	0.1

Table 4.8: Average IRR and Fleiss' Kappa values for 5 graphs with Professor 2, AA, and SMG.

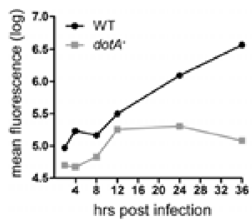
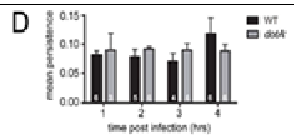
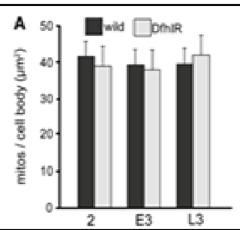
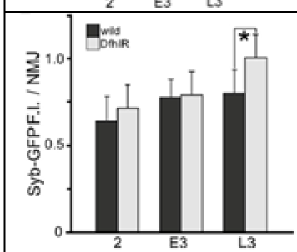
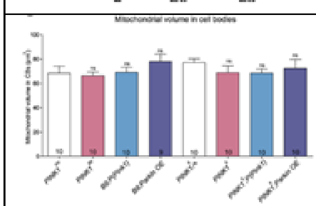
Rubric Validation with Professor2						
Graph Evaluated using the Graph Rubric	Graph Mechanics		Graph Communication		Graph Choice	
	IRR	Fleiss' Kappa	IRR	Fleiss' Kappa	IRR	Fleiss' Kappa
	0.90	0.21	1	1	0.67	-0.13
	0.90	0.22	1	1	1	1
	0.90	0.22	1	1	1	1
	0.85	0.48	1	1	1	1
	0.71	0.09	1	1	1	1

Table 4.9: Average IRR and Fleiss' Kappa values for 5 graphs with Professor 3, AA, and SMG.

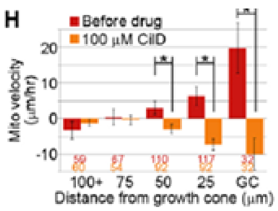
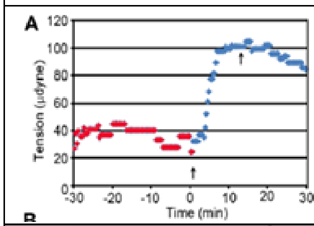
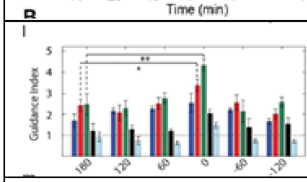
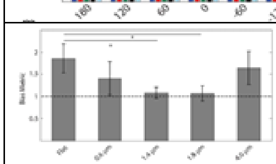
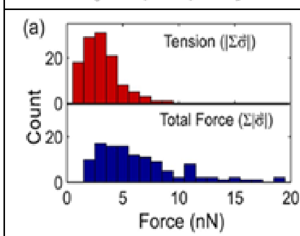
Rubric Validation with Professor 3						
Graph Evaluated using the Graph Rubric	Graph Mechanics		Graph Communication		Graph Choice	
	IRR	Fleiss' Kappa	IRR	Fleiss' Kappa	IRR	Fleiss' Kappa
	0.85	0.46	0.83	-0.2	0.77	-0.29
	0.81	0.14	0.66	-0.5	0.77	-0.29
	0.66	0.21	0.66	-0.33	0.77	-0.2
	0.71	0.30	0.66	-0.33	0.89	-0.13
	0.81	0.29	0.66	-0.33	0.89	-0.125

Table 4.10: Displayed are the values of agreement for the entire graph rubric between AA-SMG, professor (P)-AA, and P and SMG. Values in each cell are Cohen's kappa. The highlighted value in each cell indicates the highest agreement of the three pairs.

Professors	Graph 1		Graph 2		Graph 3		Graph 4		Graph 5	
1	AA-SMG	0.429	AA-SMG	0.844	AA-SMG	0.526	AA-SMG	0.667	AA-SMG	0.467
	P1-AA	0.294	P1-AA	0.422	P1-AA	0.385	P1-AA	0.091	P1-AA	0.625
	P1-SMG	-0.154	P1-SMG	0.211	P1-SMG	0.333	P1-SMG	0.143	P1-SMG	0.484
2	AA-SMG	0.25	AA-SMG	-0.091	AA-SMG	0.75	AA-SMG	0.818	AA-SMG	0.667
	AA-P2	0.429	AA-P2	1	AA-P2	0.625	AA-P2	0.455	AA-P2	0.308
	SMG-P2	-0.125	SMG-P2	-0.091	SMG-P2	0.429	SMG-P2	0.333	SMG-P2	0.238
3	AA-SMG	0.4	AA-SMG	-0.091	AA-SMG	0.222	AA-SMG	0.526	AA-SMG	0.636
	AA-P3	0.4	AA-P3	0.077	AA-P3	0.273	AA-P3	0.163	AA-P3	0.065
	SMG-P3	0.526	SMG-P3	0.273	SMG-P3	0.226	SMG-P3	0.1	SMG-P3	0.04

IRR results from the third phase of validation of textbook graphs for the graph mechanics, communication, and choice categories with AA and SMG showed a substantial agreement with value of 0.88. This indicates that critiquing textbook graphs using the graph rubric is fairly consistent. Graphs in textbooks seem to have a different focus than graphs found in other science media. For instance, textbook graphs rarely displayed a key, so the graphs did not convey information on the different colors represented by lines, bars, or dots, sample size, or the number of trials.

To account for the lack of information conveyed by different colored data points, the textbook graphs were sometimes accompanied by annotations. Graphs in textbooks did not have a title at the top of the graph, and so when evaluating this category, SMG and AA used the first sentence of the figure legend as the title, since this is the convention used by textbooks and science journals (Mack, 2013). Even then, we noticed that textbook graph titles were not descriptive because they did not mention either the subject, the variables presented by the graph, or experimental details that help in understanding the take-home message. Similarly, the graphs had vague axes labels, which sometimes made it difficult to understand the take-home message (see Figure 5.9 in Integrating Concepts in Biology for an example).

We also noted the inappropriate presence of gridlines on graphs with unscaled axes. These gridlines gave a false idea that the data presented are quantifiable, where in fact; they are sketches of ideas (See Figure 45.11 in Life the Science of Biology for an example). It was also difficult to extract the type of data plotted in the graph (See Figure 1.21 in Integrating Concepts in Biology for an example). Furthermore, it was difficult to deduce the type of data plotted since the axes labels were not descriptive and the key did not indicate the sample size or the number of trials. Textbook graphs generally exhibited appropriate graph choice, but for ease of visualization, line and symbol graphs should be constructed instead of pure line graphs (See Figure 8.9 in Life the Science of Biology for an example). If there were multiple graphs present in a figure, they were not labelled independently, so in order to interpret one graph out of a set of graphs, the graph reader had to know how to match the axes labels, how to extract information, and interpret graphs.

Another thing that textbooks have to be careful about when constructing graphs is how they make use of color and cartoon images. There were numerous graphs that had distracting background colors and cartoons, that took away the focus from the data presented on the graph (see Figure 20.3 in Discover Biology for an example).

Even though we analyzed graphs from four traditional biology paper textbooks and one from a novel, online textbook (Integrating Concepts in Biology), we saw similar features across all textbooks, which tells us that limitation of space is not an issue for data presentation. Quality of graphs in textbooks is similar to the graphs constructed by undergraduate students in our previous study in chapter 3. This analysis suggests that one reason that students are not constructing well labelled graphs is because they are not being exposed to well labelled graphs in their science classes. In order to help and teach students by example, textbook authors should focus on perfecting their graphs in all areas of graph mechanics, communication, and type of data displayed.

4.4.6 Discussion from Graph Rubric Validation

IRR results across the undergraduate students, professors, and textbook graphs showed a substantial to almost perfect agreement. However, Fleiss' kappa values ranged from -0.33 to 1. One explanation for this large range could be due to the expertise with graphing who were validating graphs. Past research has documented sixth (diSessa et al., 1991) and eighth grade (diSessa, 2002) students critiquing external representations but we are unaware of studies or curriculum practices that ask students to critique representations (graphs) with a rubric, so this exercise alone may have been a novel experience for some undergraduate students and may have caused them to not understand the reason for this exercise with using the graph rubric to evaluate graphs, which could have caused them to overlook certain aspects of graphs. Specifically for graphs 4 and 5, we noticed that undergraduate students were not bothered by the dark background or colors, indicated by 4 of the 7 students rating the aesthetics as "excellent" for graph 4, and 6 of the 7 rating the aesthetics for graph 5 as "excellent". One possible explanation for the variability in Fleiss' kappa values with the professors could be that they were provided with graphs from within their sub discipline, which could have biased their judgment. Professors, unlike AA and SMG, are familiar with the experimental contents of the graphs and understand the field so well that they may have overlooked details that the graph failed to display. Another possible explanation for the variability in kappa values can be attributed to the small sample size of professors validating the rubric. For future analysis, we plan to utilize a larger sample size, and construct a model for predictive and analytical purposes.

Although we report large variability in our kappa values, students who participated in graph rubric validation gave us favorable feedback on the graph rubric (Appendix 4I). All seven students stated that they did not encounter any problems with the rubric, the language in the rubric was easy to understand, and stated that the rubric was easy to understand. One student suggested adjusting the point values so that points are awarded depending on the severity of what the graph is missing in the category. All seven students stated that the graph rubric was helpful to them because if provided them with the key elements and explanations that needed to be included in the graph.

Additionally, all students said that they would like to use this rubric in their other science classes and would recommend their friends to also use the rubric for data presentation.

All three professors reported that they did not encounter any problems with the rubric, and the language was very easy to understand. When asked for suggestions for the rubric, one professor suggested that we allow the user to define purpose and hypothesis. Professors validating the rubric were given the purpose and hypothesis because we wanted to focus their time and attention on the graph rubric, but having students provide the purpose and hypothesis is a great strategy to implement in the classroom. We will come back to this idea in chapter 5. When asked if the professors will use the rubric in their classroom and if the students will find this useful, two of the three professors said “yes” to both questions. Since this survey is anonymous, we could not approach the professor who said “no” to both of these questions, but we think that the reason they said that this rubric would not be useful to them or their students could be due to the context of the class they are currently teaching. Maybe the subject matter does not present very many opportunities for their students to utilize this tool to inform their graph construction or analysis.

4.5 Recommended uses of the Graphing Materials

As more students are involved in data collection and analysis as part of biology curricula (Brewer & Smith 2011; Olson & Riordan 2012; NGSS Lead States 2013; College Board 2015), they will be confronted with issues revolving around graph choice and construction. The three graphing materials can be used in three different ways. First, they can be integrated throughout multi-week or semester-long science laboratory courses where students are actively collecting data and communicating their findings. Actively engaging in the step-wise process will increase students' confidence with graph choice, construction, and will enable them to critically think about their data and graphs within the classroom context and outside of the classroom. See Appendices 4B-4D for a sample scenario with dataset, filled out example of the step-by-step guide, and the resulting graphs.

By having students think about the type of data they are collecting and having them refer to the guide for data displays while they are planning their data, will broaden their knowledge with graphs and force them to think about the utility, advantages, and disadvantages of different graphs. If students are not actively engaged in data collection, instructors can provide students data from publicly-available sources (Gapminder) or from data repositories of peer-reviewed research articles to give students practice graphing data. A second suggestion is to use the guide for data displays to critique the authors' choice of graphical representations found in textbooks, primary literature, and in the popular media and link them to their claims, By having students think critically of existing graphs will encourage them to become data literate and informed citizens. A third suggestion is to utilize the graphing materials strictly as assessment tools. The instructor can give students access to the guide for data displays and quiz them periodically throughout the semester using the blank table version of this handout (Appendix 4A).

Learning gains pertaining to graph knowledge and increasing gains in MRC categories function, invention, and reflection can be formally assessed by having students fill out the blank table at the beginning and end of the semester. In order to increase grains in the fourth MRC component, critique, instructors should encourage students to use the graph rubric to critique graphs found in science and non-science media, or graphs created by their peers in the classroom. Additionally, instructors can utilize the graph rubric to provide feedback to students on their graph's mechanics, communication ability, and graph choice. In the next chapter of this dissertation, we present a novel classroom study to illustrate the usage of these graph materials and their benefits towards increasing students'graph choice, construction, and overall awareness of various types of graphs.

4.6 Critique and Conclusions

The strengths of the step-by-step guide is its ability to successfully guide students in a sequential and methodological manner from raw data to a finished graphical representation. The guide prompts a reflective approach to the process of graph creation which aids the development of MRC. Additionally, the instructor obtains instant formative feedback, which is essential for targeted instruction. The weakness of the step-by-step guide is the lack of explicit guidance to types of graphs available for use. However, this limitation can be overlooked if the step-by-step guide is paired with our guide to data displays or other similar resources (Duke et al. 2015, Graphing Tutorial, Webber et al. 2014).

Since the step-by-step guide was developed from think-aloud interviews, in which the mode of graph construction was pen and paper, we bring to light another weakness that educators and practitioners can relate to. Since graphs are constructed using software, the step-by-step guide does not incorporate Tufte's tips for displaying data in an effective and aesthetically pleasing manner, devoid of chartjunk (Tufte 1983). This limitation can be overlooked when the step-by-step guide is used with the graph rubric, which encompasses Tufte's tips for data displays. The main strength of the guide to data displays is the organized manner in which information on graph usage, advantages, and disadvantages is presented. Consequently, one weakness is that only six graphs are represented. However, since extensive research was done looking at graphs from textbooks and primary literature to decide what graphs to display, we think that this is acceptable for beginning students. Instructors may use this table to expand to other graph types as they see fit for the students at more advanced levels of learning.

The strengths of the graph rubric are in its ability to quickly allow the instructor to provide the students with quick, unbiased, and targeted feedback. By utilizing this rubric in the classroom and exposing it to students not only communicates the learning objectives but also the expectations of a well-constructed graph. The graph has three different levels of achievement (excellent, needs improvement, and unsatisfactory) and provides the instructor with freedom to assign points to each category however they choose. Although the graph rubric emerged from existing graphing literature and ideas from our own research in both clinical (chapter 2) and naturalistic settings (chapter 3), after validating it with graphs from textbooks and research articles, we realized that research articles feature graphs that are data dense and more complex than graphs that students make in the classroom and graphs present in introductory biology textbooks. Because of the limited amount of space allotted for figures by research journals, graphs are usually small and do not have titles, labels and key. All of this information is found in the figure legend, a category not present in our graph rubric. Although figure legends are informative accompaniments to a graph, they were not part of the graph rubric. Here we present four reasons for excluding figure legends and suggestions for future research aimed at incorporating figure legends into the classroom. The first reason is that writing figure legends is an art that requires knowledge on which methods and results to include, and the message that needs to be conveyed to the audience (Rodriguez, 2013). In order to understand how a figure legend is written, think-aloud interviews should be conducted between expert and novice biologists, similar to those in chapter 2, but focused on figure legends.

Secondly, figure legends may be more useful for figures that are not meant to be stand alone representations, such as microscopic images, photographs, diagrams, or figures with multiple components. Since graphs are meant to be stand-alone representations with the purpose of conveying complex data in a quick and efficient manner, it is recommended that graphs are properly labelled (Mack, 2013). To test the necessity of figure legends, in think-aloud interviews, participants can be presented with various images and asked to interpret them. We can predict that a well-labeled graph will be the easiest to interpret as compared to some of the other types of figures normally found in science journals or textbooks.

The third reason is that the contexts that informed the development of the graph rubric (think-aloud interviews, chapter 2 and classroom oral PowerPoint presentations, chapter 3) were from verbal data, where participants were not instructed to write a figure legend. Future work can focus exclusively on figure legends and see how they are depicted in science journals, textbooks, and student lab reports.

The fourth reason for not including figure legends in the graph rubric goes back to the idea of general science literacy. As science educators, one of our goals is to ensure that when students encounter data in the real world, that they can use their scientific knowledge to draw their own evidence-based conclusions (Gormally et al., 2012). Sometimes figure legends can bias the reader's interpretation to the intended purpose by the author. Furthermore, since most graphs found in non-science media are not accompanied with figure legends, it is necessary for the student to be able to analyze the graph based on the data and graph mechanics and draw their own conclusions.

Graphs constructed by students in the physiology class are different from graphs in journals because students are instructed to be more explicit whereas the graphs in journals elude to being more implicit. For a novice reader, graphs combined with a figure legend can be a challenging feat, which allows us to recommend that graphs seen in research journals should follow graphing conventions mentioned in the rubric so that they can be accessible to students who are learning to read scientific literature. Furthermore, there has been a push to improve how graphs are displayed in research journals (Rougier et al., 2014; Weissgerber et al 2015).

4.7 Summary

In this chapter, we presented the purpose, development, validation, and usage of three graphing materials: Guide to data displays, step-by-step guide to data communication, and the graph rubric. The development of all three graphing materials was informed from previous literature on graphing with students at the K-12 and undergraduate levels and from our expert-novice study (chapter 2) and naturalistic classroom study (chapter 3). The purpose of the guide to data displays is to increase students' MRC by exposing them to different types of graphs, their usage, advantages, and disadvantages. The purpose of the step-by-step guide is to provide students with a framework for data presentation, with encompasses all parts of the MRC (Table 4.1). Lastly, the purpose of the graph rubric is to help instructors provide quick and consistent feedback on students' graphs and for students to use when constructing and critiquing graphs.

All three materials were validated by expert and novice biologists and the usage of all three materials in an upper-level physiology classroom at a Midwestern University showed improved student knowledge with graph choice, construction, and alignment. Future work can focus on the utility of these materials at the K-12 to see if they help remediate some of the common difficulties previously reported with graphing.

Acknowledgments

We thank all of the biology undergraduate students, graduate students, and professors who have participated in all of our studies so far. We also thank our research group, PIBERG, for their early feedback on this project. We also thank Mike Melomed for helping us with statistics. This project emerged from ideas initiated within the Biology Scholars Research Residency program.

References

Ainley, J. (2000). Transparency in graphs and graphing tasks: An iterative design process. *The Journal of Mathematical Behavior*, *19*(3), 365-384.

Allen, D., & Tanner, K. (2006). Rubrics: tools for making learning goals and evaluation criteria explicit for both teachers and learners. CBE-Life Sciences Education, 5(3), 197-203.

Angra, A., Gardner, S.M. (2014). *Using Student Laboratory Reflections to Reveal Reasoning behind Graph Choice and Construction*. In: Purdue University Chapter The Society of Sigma Xi.

Angra, A. and Gardner, S.M. (2016). Evaluating Meta-Representations Competence with Graph Choice and Construction along the Novice-Expert Continuum in Biology. Manuscript submitted for publication in CBE-LSE.

Angra, A. and Gardner, S.M. (2015). *Development of an Analytic Rubric to Evaluate Undergraduate Student Graphs and Diagnose Difficulties in Graph Choice and Construction*. In: Society for the Advancement of Biology Education Research (SABER).

Berg, C. A., & Phillips, D. G. (1994). An investigation of the relationship between logical thinking structures and the ability to construct and interpret line graphs. *Journal of Research in Science Teaching*, *31*(4), 323-344.

Bertin, J. (1983). Semiology of graphics: diagrams, networks, maps. Madison, WI: Univ Wisconsin Press.

Bowen, G. M., & Roth, W. M. (2005). Data and graph interpretation practices among preservice science teachers. *Journal of research in science teaching*,*42*(10), 1063-1088.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (1999). *How people learn: Brain, mind, experience, and school*. National Academy Press.

Brasell, H. M., & Rowe, M. B. (1993). Graphing skills among high school physics students. *School Science and Mathematics*, *93*(2), 63-70.

Bray-Speth, E. B., 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(3), 323-332.

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

Bruno, A., & Espinel, M. C. (2009). Construction and evaluation of histograms in teacher training. *International Journal of Mathematical Education in Science and Technology*, *40*(4), 473-493.

Campbell, A., Heyer, L., Paradise, C., Sellers, P., & Barsoum, M. (2011, January). Integrating Concepts in Biology for Introductory College Biology. In *Molecular Biology of the Cell* (Vol. 22). American Society of Cell Biology.

Chi, M. T., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices*. *Cognitive science*, *5*(2), 121-152.

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*(1), 13-19.

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*(3), 317-322.

College Board. (2015). *AP Biology Course and Exam Description.* Retrieved from https://secure-media.collegeboard.org/digitalServices/pdf/ap/ap-biology-course-and-exam-description.pdf

College Board. (2012). AP Biology Quantitative Skills: A Guide for Teachers. Retrieved from http://apcentral.collegeboard.com/apc/public/repository/AP_Bio_Quantitative_Skills_Guide-2012.pdf

Cumming, G., Fidler, F., & Vaux, D. L. (2007). Error bars in experimental biology. *The Journal of cell biology*, *177*(1), 7-11.

DiFazio, M. H. (1990). Graphics Software Side by Side. The Mathematics Teacher, 83(6), 436-446.

diSessa, A.A., Hammer, D., Sherin, B. (1991). Inventing graphing: Meta-representational expertise in children. Journal of Mathematical Behavior, 10(2), 117-160.

diSessa, A.A., & Sherin, B. L. (2000). Meta-representation: An introduction. *The Journal of Mathematical Behavior*, *19*(4), 385-398.

diSessa, A. A. (2002). Students' criteria for representational adequacy. *Symbolizing, modeling and tool use in mathematics education*, 105-129.

diSessa, A. A. (2004). Metarepresentation: Native competence and targets for instruction. *Cognition and instruction*, *22*(3), 293-331.

Drummond, G. B., & Tom, B. D. (2011). Presenting data: can you follow a recipe?. *Clinical and Experimental Pharmacology and Physiology*, *38*(12), 787-790.

Drummond, G. B., & Vowler, S. L. (2011). Show the data, don't conceal them. *Advances in physiology education*, *35*(2), 130-132.

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*(22), 3040-3059.

Duncan, D. B., Lubman, A., & Hoskins, S. G. (2011). Introductory biology textbooks under-represent scientific process. Journal of Microbiology & Biology Education: JMBE, 12(2), 143.

Ehrenberg, A. S. C. (1977). Rudiments of numeracy. *Journal of the Royal Statistical Society. Series A (General)*, 277-297.

Elliott, A.C., Hynan L.S., Reisch, J.s., & Smith, J.P. (2006). Preparing Data for Analysis Using Microsoft Excel. *Journal of Investigative Medicine*, 54(6), 334-341.

Few, S. (2004). *Show me the numbers: Designing tables and graphs to enlighten* (Vol. 1, No. 1). Oakland, CA: Analytics Press.

Fleiss, J. L. (1981). Statistical methods for rates and proportions. John Wiley & Sons.

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. *The Journal of hand surgery*, *37*(3), 591-596.

Gormally, C., Brickman, P., & Lutz, M. (2012). Developing a Test of Scientific Literacy Skills (TOSLS): measuring undergraduates' evaluation of scientific information and arguments. *CBE-Life Sciences Education*, *11*(4), 364-377.

Gapminder. *Data in Gapminder World*. Retrieved from http://www.gapminder.org/data/

Graphing Tutorial. *Graphing Tutorial*. Retrieved from http://nces.ed.gov/nceskids/help/user_guide/graph/index.asp

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

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. biometrics, 159-174.

Olson, S., & Riordan, D. G. (2012). Engage to Excel: Producing One Million Additional College Graduates with Degrees in Science, Technology, Engineering, and Mathematics. Report to the President. *Executive Office of the President*.

Humphrey, P. B., Taylor, S., & Mittag, K. C. (2014). Developing Consistency in the Terminology and Display of Bar Graphs and Histograms. *Teaching Statistics*, *36*(3), 70-75.

IBM Corp. Released. (2013). IBM SPSS Statistics for Windows, Version 22.0. Armonk, NY: IBM Corp.

Interactive Statistics Map. *Numb3R5 Count! Interactive Statistics Map*. Retrieved from http://bioquest.org/numberscount/statistics-map-2/

Koedinger, K. R., & Anderson, J. R. (1990). Abstract planning and perceptual chunks: Elements of expertise in geometry. *Cognitive Science*,*14*(4), 511-550.

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

Kuhn, D., & Udell, W. (2003). The development of argument skills. *Child development*, *74*(5), 1245-1260.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. biometrics, 159-174.

Leonard, J. G., & Patterson, T. F. (2004). Simple computer graphing assignment becomes a lesson in critical thinking. *NACTA journal*, *48*(2), 17-21.

Livescribe, Inc. (2006). *Livescribe Smartpens*. Retrieved from http://www.livescribe.com/en-us/smartpen/

Mack, C. (2013). How to write a good scientific paper: Figures, part 1.*Journal of Micro/Nanolithography, MEMS, and MOEMS*, *12*(4), 040101-040101.

McFarland, J. (2010). Teaching and Assessing Graphing Using Active Learning. *MathAMATYC Educator*, *1*(2), 32-39.

McHugh, M. L. (2012). Interrater reliability: the kappa statistic. Biochemia medica, 22(3), 276-282.

NGSS Lead States. (2013) *Next Generation Science Standards: For States, By States* Retrieved from http://www.nextgenscience.org/next-generation-science-standards

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*(1), 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*(6), 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(1), 95-115.

Patterson Jr, T. F., & Leonard, J. G. E-Graphing: Perfection, Paradox or Persuasion. URL http://www.uvm.edu/~jpetrill/Paterson%20%26%20Leonard%20-%20E-Graphing.pdf

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

Puhan, M. A., ter Riet, G., Eichler, K., Steurer, J., & Bachmann, L. M. (2006). More medical journals should inform their contributors about three key principles of graph construction. *Journal of clinical epidemiology*, *59*(10), 1017-e1.

Quinn, G. P., & Keough, M. J. (2002). Experimental design and data analysis for biologists. Cambridge University Press.

Raven, P. H., Johnson, G. B., Losos, J. B., Mason, K. A. & Singer, S. R. (2008): Biology, 8th ed.; Boston (McGraw-Hill).

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Rodrigues, V. (2013). Tips on effective use of tables and figures in research papers. *Editage Insights (04-11-2013)*.

Roth, W. M., Bowen, G. M., & McGinn, M. K. (1999). Differences in graph-related practices between high school biology textbooks and scientific ecology journals. *Journal of research in science teaching*, *36*(9), 977-1019.

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

Rougier, N. P., Droettboom, M., & Bourne, P. E. (2014). Ten simple rules for better figures. *PLOS Comput Biol*, *10*(9), e1003833.

Rybarczyk, B. (2011). Visual Literacy in Biology: A Comparison of Visual Representations in Textbooks and Journal Articles. Journal of College Science Teaching, 41(1), 106-114.

Sadava, D. E., Hillis, D. M., Heller, H. C., & Berenbaum, M. (2009). Life: the science of biology (Vol. 2). Macmillan.

Saxon, E. (2015). Beyond bar charts. *BMC biology*, *13*(1), 60.

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*(1), 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*(6), 750-756.

Singh-Cundy, A., Shin, G. (2010). Discover biology (6th Edition). Sinauer Associates, Inc.

Tairab, H. H., & Khalaf Al-Naqbi, A. K. (2004). How do secondary school science students interpret and construct scientific graphs?. *Journal of Biological Education*, *38*(3), 127-132.

Tanner, K. D. (2012). Promoting student metacognition. *CBE-Life Sciences Education*, *11*(2), 113-120.

Tufte, E. R., & Graves-Morris, P. R. (1983). *The visual display of quantitative information* (Vol. 2, No. 9). Cheshire, CT: Graphics press.

Urry, L. A., Cain, M. L., Wasserman, S. A., Minorsky, P. V., Jackson, R. B., & Reece, J. B. (2014). *Campbell biology in focus*. Pearson.

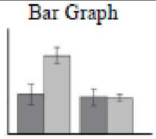
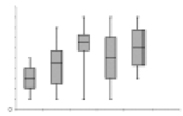
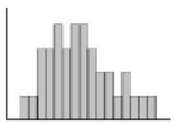
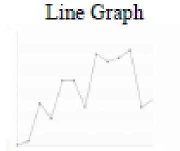
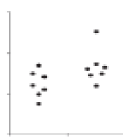
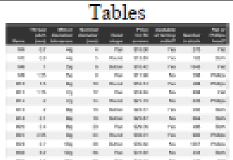
Viera, A. J., & Garrett, J. M. (2005). Understanding interobserver agreement: the kappa statistic. Fam Med, 37(5), 360-363.

Webber, H., Nelson, S. J., Weatherbee, R., Zoellick, B., Schauffler, M.(2014). The Graph Choice Chart. *The Science Teacher 81*, 37-43.

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 Biol*, *13*(4), e1002128.

Wolf, K., & Stevens, E. (2007). The Role of Rubrics in Advancing and Assessing Student Learning. *Journal of Effective Teaching*, *7*(1), 3-14.

Appendix 4A: Blank Guide to Data Displays

Data Display Type	Usage	Advantages	Disadvantages
Bar Graph 			
Box and Whisker Plot 			
Histogram 			
Line Graph 			
Dot Plot 			
Scatterplot 			
Tables 			

Appendix 4B: Graph Rubric for Student Peer Evaluations

Directions: Use the following rubric to evaluate and provide constructive feedback to your classmates. Please write comments in the boxes. You will earn full points for thoughtfully completing this assignment.

Research Question:

Hypothesis:

	ID of graph or set of graphs:	**Present and Appropriate**	**Present but Needs Improvement**	**Absent**
Graph Mechanics	**Descriptive title** • P/A-Should be: a) in the form of a statement, b) mention the subject, c) appropriate variables, and d) include relevant details about the experiment that help understand the take home message. • NI- If the title is missing any one of the four points mentioned above.			
	Label for the X axis (e.g. time) • P/A- Should be appropriate and descriptive for the experiment. For graphs with categorical independent variables, there needs to be a label under each set of data and a larger label under all data plotted. • NI- If the label is missing any one of the points mentioned above.			
	Label for the Y axis (e.g. heart rate) • P/A- Should be appropriate and descriptive for the experiment. If the data is manipulated (average, change, percentage, etc.), then it should be indicated on the y axis. • NI- If the label is missing any one of the points mentioned above.			
	Units for the X axis (e.g. seconds) • P/A- Should be appropriate and descriptive for the data displayed. • NI- If the units are not appropriate or descriptive.			
	Units for the Y axis (e.g. average beats per minute) • P/A- Should be appropriate and descriptive for the data displayed. • NI- If the units are not appropriate or descriptive.			
	Scale (appropriate intervals and range for data) • P/A- Should be appropriate for the data displayed such that the increments are clear and without clutter and includes appropriate significant figures. If the scale is discontinuous or doesn't start at the origin, it should be indicated by a break in the axis. • NI- If the scale is not appropriate for the data such that it is cluttered, does not include appropriate significant figures, and/or if the scale does not indicate axis break.			
	Key (defines different data sets that are plotted) • P/A- Should be appropriate and descriptive for the data displayed. It should include: a) descriptions of different colors (if applicable), b) the sample size and c) the number of trials. • NI- If the key is not descriptive and does not indicate the sample size.			

		Excellent (E)	**Needs Improvement (NI)**	**Unsatisfactory (U)**
Communication	**Ease of Understanding-Aesthetics** • E- If the graph is aesthetically pleasing, meaning that: a) the data plotted takes up sufficient room in the Cartesian plane, b) makes use of legible size font, c) the x and y axis lines are clear and legible, d) the graph displays data in an appropriate number of bars and lines, and e) is devoid of chart junk elements such as: distracting background colors, patterns, and dark gridlines • NI- If the graph has one of the following flaws: a) the graph displays too much white space, b) the font size is too small, c) the x and y axis lines are not clear and legible, d) the graph shows too many bars or lines OR e) elements of chart junk are clouding interpretation of data. • U- If the graph has multiple flaws, which interfere with the understanding and interpretation of data.			
	Ease of Understanding-Take home message • E- If the graph has sound construction and mechanics that allow for clear sorting of trends and take home message. • NI- If data trends are difficult to observe or it is difficult to formulate a proper take home message. • U- If the graph is ineffective at communicating data trends and take home message, such that it causes confusion.			
Graph Choice	**Graph Type (Bar, line, scatter, dot, box and whisker)** • E- If data displayed in a graph is appropriate for both independent and dependent experimental variables (i.e. categorical and continuous) and data. (*Referring to the data form) • NI- If data displayed in a graph is a) not suitable for either the dependent or independent experimental variables OR b) there is a better way to present data. • U- If the graph type is not suitable for both experimental variables.			
	Data Displayed (Raw, Averages, Changes, Percentage) • E- If the graph indicates the type of data (ex. Raw, averages, etc) that are plotted. There should be a clear distinction between raw data and manipulated data based on the information presented in the key (ie, sample size and number of trials) and axis label. If the graph is showing averages, then it should also be accompanied with STDEV or error bars. • NI- If the graph is missing one of points mentioned above • U- If data type is inappropriate for the graph type			
	Alignment *(at least one of the graphs presented should align with the research question and hypothesis. Other graphs can be exploratory.)* • E- If the graph is completely aligned with the research question and/or hypothesis. In other words, the independent, dependent variables, and information about the experiment are explicit. • NI- If the graph is partially aligned with the research question and/or hypothesis. In other words, the graph is missing information about either the independent, dependent, or details about the experiment. • U- If the graph is not aligned with the research question and/or hypothesis.			

What are some advantages of this representation?

What are some disadvantages of this representation?

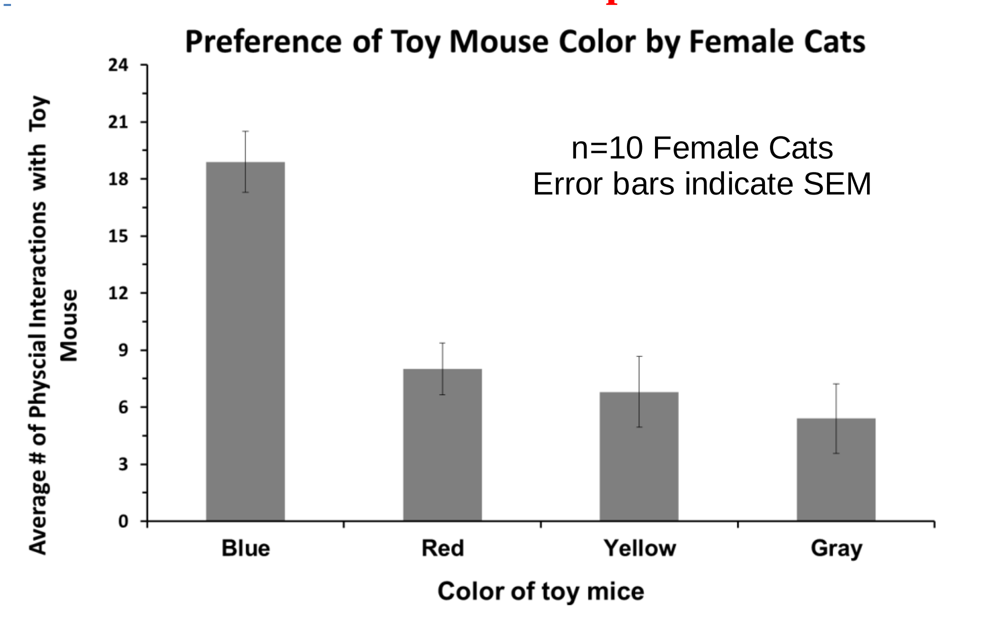
Appendix 4C: Sample Dataset for Classroom Implementation

Imagine you are a toy designer for a cat toy company, who frequently collaborates with scientists who study animal vision. You know that cats are crepuscular creatures, meaning they are active during dawn and dusk, and have excellent night vision because their eyes have up to eight times more rod cells than humans. Unfortunately, the same cannot be said about their color vision. Cats have 12 times fewer cone cells than humans, which limits the number of colors they can see. Recent studies have shown that cats can discriminate between blue and gray colors really well. Scientists still are not sure if cats can discriminate between other colors. Your job is to figure out what color mice toys cats physically interact with the most. Given the research conducted, you hypothesize to find a difference in cat color preference. Specifically, you predict that cats will have the most physical interactions with the blue colored mouse over other colors. You set up a small experiment to collect preliminary data. You will determine their color preference by videotaping each cat separately over a span of 10 hours. Each cat will be presented with 4 different colored mice in a 10’ x 15’ room. You test out the mice on 10 different young female cats between the ages of 1-2 years and you display your data in the table below.

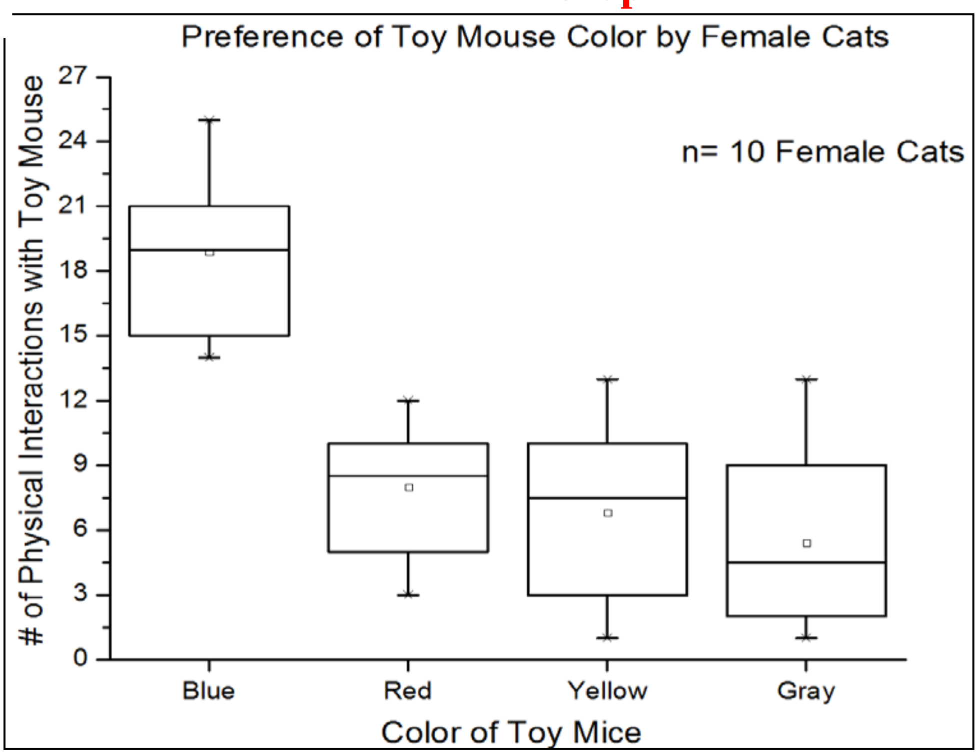
	Number of Physical Interactions with the Colored Toy Mouse			
Cat #	Blue	Red	Yellow	Gray
1	19	11	7	2
2	18	8	6	5
3	15	10	9	10
4	15	12	13	13
5	20	5	1	6
6	25	4	3	1
7	23	3	4	1
8	14	10	9	9
9	21	8	10	4
10	19	9	11	3

Appendix 4D: Sample Graphs from Dataset

Graph 1

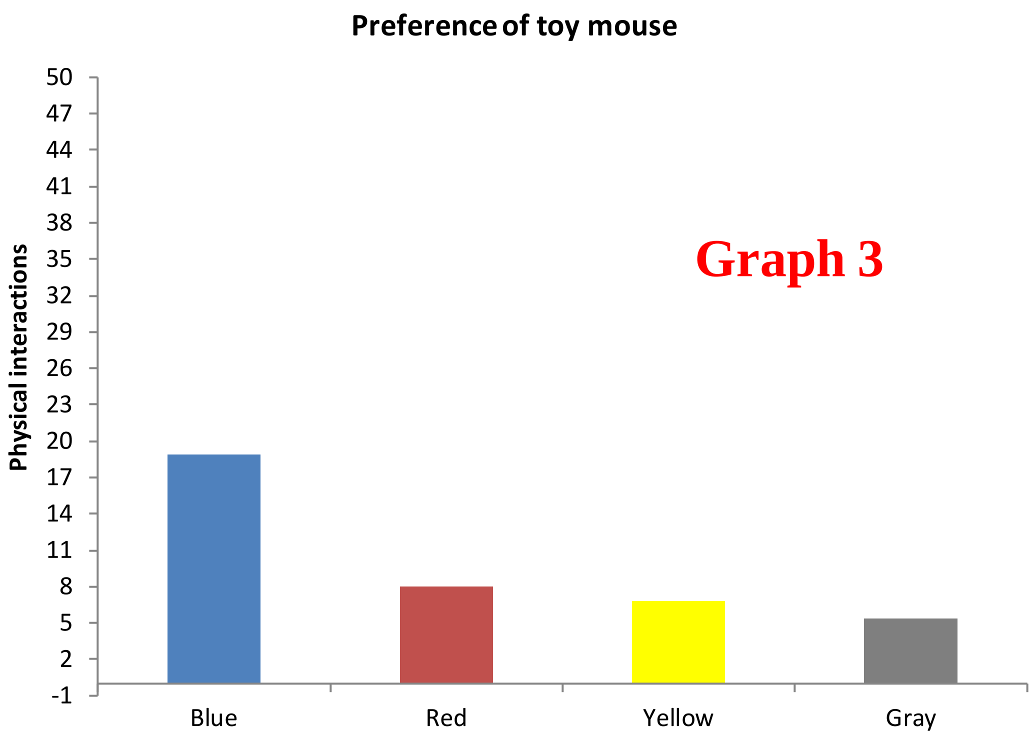


Graph 2



Appendix 4D continued

Appendix 4D continued



Appendix 4E: Filled out Step-by-Step Guide

Phases 1-3	Elements	Notes for your Experiment
1. Planning- In this phase, you must organize your data and decide on the message you want to communicate in your graph. It helps to first conceptualize the whole task before executing it.	Step 1- Revisit your research question and hypothesis and ask yourself, what is it that you want the graph to show?	RQ- Do young female cats display a strong color preference for toy mice? HYP-There is a difference in color preference amongst young female cats. PRED-Since cats can discriminate between blue and black colors, toy mice of these colors will show the most interaction time.
	Step 2- Identify your independent and dependent variables.	Independent variable- Color of toy mice, specifically, blue, red, yellow, and gray. Dependent variable- Physical interactions with the toy mouse.
	Step 3- Classify your variables as either categorical or continuous.	Color of mice, specifically blue, red, yellow, and gray are categorical variables. Number of physical interactions with the toy mouse is a continuous variable.
	Step 4- Decide whether or not you need to manipulate your data.	I will take an average and calculate the standard error of the number of hours played with each colored mouse. I would also like to see the variability in interaction time for each color mouse among the 10 cats.
	Step 5-Decide on a graph type that will best represent your data.	I will make a **bar graph** because I would like to compare the average number of physical interactions with the different colored mice. I will make a **box plot** because I would like to compare the variability of all data points.
2. Execution- In this phase, you will actively construct a graph.	Step 6-Label the axes with your variables.	X axis- independent variable- Color of toy mice (as my large axis label); blue, red, yellow, gray (as labels for bars). Y axis (bar graph)- dependent variable- Average physical interactions with the toy mouse. Y axis (box plot)- Number of physical interactions with the toy mouse
	Step 7-Add units to the axes, if necessary.	X axis- since I have categorical data, I do not need units. Y axis- Separate units are unnecessary since they are indirectly embedded into the axis label (i.e. number of OR average number of physical interactions)
	Step 8- Adjust the scale of axes into appropriate increments for the data.	For the bar graph, since the lowest average interaction is 5.3 for the gray color and the highest is 19.5 for the blue color; I will set the scale from 0-24 with increments of 3. For the box plot, since the highest number is 25 and the lowest number is 0, I will set the scale from 0- 27 in increments of 3.
	Step 9-Include a key, if appropriate.	On each graph, I will mention that the sample size is 10 female cats. On the bar graph, I will specify the error bars as SE so the reader can differentiate them between other types of error bars like standard deviation, range, or confidence intervals [13].
	Step 10-If you are displaying the graph in a report, include a figure legend.	N/A
	Step 11-Include a descriptive title.	Preference of toy mouse color by female cats.
3. Reflection and Explanation Phase- In this phase, you will critically reflect on your graph choice, interpret your graph, and explain your answers to questions posed in steps 13-16.	Step 12-Check the alignment of your representation with your research question and hypothesis.	Yes, both graphs align.
	Step 13-What are the advantages of the representation?	The graph communicates a clear take home message; the independent (x axis) and dependent (y axis) variables are appropriate for both graphs; the bar graph is displaying averages and standard error, which are suitable for the type of data presented in a bar graph; the box plot is showing and comparing the mean, median, and distribution of the data in quartiles and whiskers; the graphs align with the research question and hypothesis.
	Step 14- What are the disadvantages of the representation?	Although the graphs are well made and align with the research question and hypothesis, a minor disadvantage would be that data for each individual female cat is not visible and properly interpreting these graphs may be difficult for the reader. Specifically, the bar graph is displaying the standard error bars, which convey different information than other types of error bars such as standard deviation, range, and confidence intervals (Cumming et al. 2007). A disadvantage of the box plot would be the unfamiliarity with the different components, ie. mean, median, quartiles, and whiskers.
	Step 15- What is the take-home message of the representation?	On average, the 10 female cats had three times more physical interactions with the blue mouse, as compared to the red, yellow, and gray mice.
	Step 16-What are some other ways that you could have represented your data?	If I was interested in the proportion of time that cats played with each mouse, I could construct a pie chart. If I was interested in individual color preferences for each female cat, I could make a stacked bar graph, with the cats on the x axis, and display the raw data (hours played) in color coded bars.

Appendix 4F: Filled out Graph Rubric

Research Question: Do young female cats display a strong color preference for toy mice?
Hypothesis: There is a difference in color preference amongst young female cats.

		GRAPH 2-BOX			GRAPH 3-BAR		
		P/A	**NI**	**X/I**	**P**	**NI**	**X/I**
Graph Mechanics	**Descriptive title** • P/A-Should be: a) in the form of a statement, b) mention the subject, c) appropriate variables, and d) include relevant details about the experiment that help understand the take home message. • NI- If the title is missing any one of the four points mentioned above.	x				**Missing (b), (c), and (d)**	
	Label for the X axis (e.g. time) • P/A- Should be appropriate and descriptive for the experiment. For graphs with categorical independent variables, there needs to be a label under each set of data and a larger label under all data plotted. • NI- If the label is missing any one of the points mentioned above.	x				**Missing larger label**	
	Label for the Y axis (e.g. heart rate) • P/A- Should be appropriate and descriptive for the experiment. If the data is manipulated (average, change, percentage, etc.), then it should be indicated on the y axis. • NI- If the label is missing any one of the points mentioned above.	x				**Not descriptive. Physical interactions with what?**	
	Units for the X axis (e.g. seconds) • P/A- Should be appropriate and descriptive for the data displayed. • NI- If the units are not appropriate or descriptive.	n/a			n/a		
	Units for the Y axis (e.g. average beats per minute) • P/A- Should be appropriate and descriptive for the data displayed. • NI- If the units are not appropriate or descriptive.	x				**since title is not descriptive, might need units.**	
	Scale (appropriate intervals and range for data) • P/A- Should be appropriate for the data displayed such that the increments are clear and without clutter and includes appropriate significant figures. If the scale is discontinuous or doesn't start at the origin, it should be indicated by a break in the axis. • NI- If the scale is not appropriate for the data such that it is cluttered, does not include appropriate significant figures, and/or if the scale does not indicate axis break.	x				**Scale is not appropriate.**	
	Key (defines different data sets that are plotted) • P/A- Should be appropriate and descriptive for the data displayed. It should include: a) descriptions of different colors (if applicable), b) the sample size and c) the number of trials. • NI- If the key is not descriptive and does not indicate the sample size.	x					**Missing key and sample size.**

Appendix 4F continued

		GRAPH 2-BOX			GRAPH 3-BAR		
		E	NI	U	E	NI	U
Communication	**Ease of Understanding-Aesthetics** • E- If the graph is aesthetically pleasing, meaning that: a) the data plotted takes up sufficient room in the Cartesian plane, b) makes use of legible size font, c) the x and y axis lines are clear and legible, d) the graph displays data in an appropriate number of bars and lines, and e) is devoid of chart junk elements such as: distracting background colors, patterns, and dark gridlines • NI- If the graph has one of the following flaws: a) the graph displays too much white space, b) the font size is too small, c) the x and y axis lines are not clear and legible, d) the graph shows too many bars or lines OR e) elements of chart junk are clouding interpretation of data. • U- If the graph has multiple flaws, which interfere with the understanding and interpretation of data.	x				Too much white space. Colored bars are distracting.	
	Ease of Understanding-Take home message • E- If the graph has sound construction and mechanics that allow for clear sorting of trends and take home message. • NI- If data trends are difficult to observe or it is difficult to formulate a proper take home message. • U- If the graph is ineffective at communicating data trends and take home message, such that it causes confusion.	x				Missing detailed info. on the experiment. Take home message is difficult to formulate.	
Graph Choice	**Graph Type (Bar, line, scatter, dot, box and whisker)** • E- If data displayed in a graph is appropriate for both independent and dependent experimental variables (i.e. categorical and continuous) and data. (*Referring to the data form) • NI- If data displayed in a graph is a) not suitable for either the dependent or independent experimental variables OR b) there is a better way to present data. • U- If the graph type is not suitable for both experimental variables.	x			x		
	Data Displayed (Raw, Averages, Changes, Percentage) • E- If the graph indicates the type of data (ex. Raw, averages, etc.) that are plotted. There should be a clear distinction between raw data and manipulated data based on the information presented in the key (ie. sample size and number of trials) and axis label. If the graph is showing averages, then it should also be accompanied with STDEV or error bars. • NI- If the graph is missing one of points mentioned above. • U- If data type is inappropriate for the graph type	x				Not sure. Missing key and lack of descriptive labels makes it difficult to deduce the type of data displayed.	
	Alignment *(at least one of the graphs presented should align with the research question and hypothesis. Other graphs can be exploratory.)* • E- If the graph is completely aligned with the research question and/or hypothesis. In other words, the independent, dependent variables, and information about the experiment are explicit. • NI- If the graph is partially aligned with the research question and/or hypothesis. In other words, the graph is missing information about either the independent, dependent, or details about the experiment. • U- If the graph is not aligned with the research question and/or hypothesis.	x					No mention of the subject or treatments in the graph.

Appendix 4G: Example of IRR between 7 Students, AA, and SMG

Graph Rubric Categories	SMG	AA	Student2	Student3	Student4	Student5	Student9	Student10	Student11	IRR
Title	1	1	2	1	1	1	1	2	1	0.77
X axis Label	2	2	2	2	2	2	2	2	2	1
Y axis Label	2	2	2	2	2	1	2	2	1	0.88
X axis Units	2	2	2	0	2	2	2	2	2	0.88
Y axis Units	2	2	2	2	2	2	2	2	2	1
Scale	2	2	2	2	2	2	2	2	2	1
Key	1	1	1	2	1	1	1	1	1	0.88
					AVG IRR Score for Graph Mechanics					**0.916**
Aesthetics	1	1	1	2	2	2	2	2	2	0.33
THM	2	2	1	2	2	2	1	2	1	0.77
Graph Type	1	1	1	2	2	2	2	2	2	0.33
Data Displayed	1	1	1	2	1	1	1	1	1	0.88
Alignment	2	2	2	2	2	2	1	1	1	0.66
			AVG IRR Score for Graph Communication & Choice							**59.4**

Fleiss Kappa for 9 raters = 0.5810 SE = 0.0596
95%CI = 0.4643 to 0.6978

Areas of disagreement

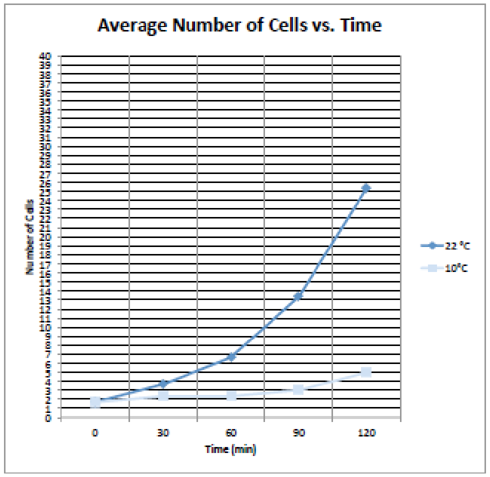
Fleiss Kappa for 9 raters = 0.1000 SE = 0.0745
95%CI = -0.0461 to 0.2461

Appendix 4H: Sample Graph 1

Researcher's Name:
Lab Section and Year:
Group Number:

Research Question: Does temperature affect the growth rate of this type of bacteria?
Hypothesis: The bacteria will have different growth rates at different temperatures.

	ID of graph or set of graphs:	**Present/ Appropriate**	**Present but Needs Improvement**	**Absent**
Graph Mechanics	**Descriptive title** P/A-Should be: a) in the form of a statement, b) mention the subject, c) appropriate variables, and d) include relevant details about the experiment that help understand the take home message. NI- If the title is missing any one of the four points mentioned above.		Needs to be in a statement, mention the subject (bacteria), and include details about the experiment (temperatures)	
	Label for the X axis (e.g. time) P/A- Should be appropriate and descriptive for the experiment. For graphs with categorical independent variables, there needs to be a label under each set of data and a larger label under all data plotted. NI- If the label is missing any one of the points mentioned above.	x		
	Label for the Y axis (e.g. heart rate) P/A- Should be appropriate and descriptive for the experiment. If the data is manipulated (average, change, percentage, etc.), then it should be indicated on the y axis. NI- If the label is missing any one of the points mentioned above.		Should say "Average Number of Cells", since that is the data plotted.	
	Units for the X axis (e.g. seconds) P/A- Should be appropriate and descriptive for the data displayed. NI- If the units are not appropriate or descriptive.	x		
	Units for the Y axis (e.g. average beats per minute) P/A- Should be appropriate and descriptive for the data displayed. NI- If the units are not appropriate or descriptive.	x		
	Scale (appropriate intervals and range for data) P/A- Should be appropriate for the data displayed such that the increments are clear and without clutter and includes appropriate significant figures. If the scale is discontinuous or doesn't start at the origin, it should be indicated by a break in the axis. NI- If the scale is not appropriate for the data such that it is cluttered, does not include appropriate significant figures, and/or if the scale does not indicate axis break.		The y axis can be re-scaled so that it goes from -2 to 26 and a proper increment should be chosen so that numbers do not clutter the y axis.	
	Key (defines different data sets that are plotted) P/A- Should be appropriate and descriptive for the data displayed. It should include: a) descriptions of different colors (if applicable), b) the sample size and c) the number of trials. NI- If the key is not descriptive and does not indicate the sample size.		Since the graph is showing averages, a key showing the sample size for each temperature is necessary.	



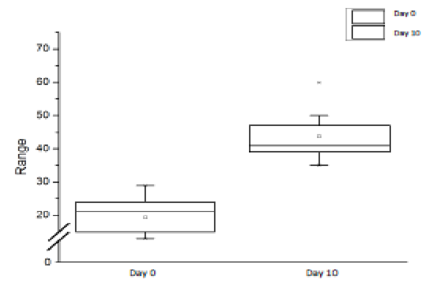
Appendix 4H continued

		Excellent (2)	Needs Improvement (1)	Unsatisfactory (0)
Communication	**Ease of Understanding-Aesthetics** **E-** If the graph is aesthetically pleasing, meaning that: a) the data plotted takes up sufficient room in the Cartesian plane, b) makes use of legible size font, c) the x and y axis lines are clear and legible, d) the graph displays data in an appropriate number of bars and lines, and e) is devoid of chart junk elements such as: distracting background colors, patterns, and dark gridlines **NI-** If the graph has one of the following flaws: a) the graph displays too much white space, b) the font size is too small, c) the x and y axis lines are not clear and legible, d) the graph shows too many bars or lines OR e) elements of chart junk are clouding interpretation of data. **U-** If the graph has multiple flaws, which interfere with the understanding and interpretation of data.			This graph has multiple flaws: the gridlines are distracting, there is unnecessary white space above the data lines, and the two data lines are similar in color. Overall, there is a lot of chartjunk in this graph.
	Ease of Understanding-Take home message **E-** If the graph has sound construction and mechanics that allow for clear sorting of trends and take home message. **NI-** If data trends are difficult to observe or it is difficult to formulate a proper take home message. **U-** If the graph is ineffective at communicating data trends and take home message, such that it causes confusion.		Although it is easy to note the trends in this graph, it is difficult to formulate an accurate take home message since the subject, bacteria is not mentioned in the graph.	
Graph Choice	**Graph Type (Bar, line, scatter, dot, box and whisker)** **E-** If data displayed in a graph is appropriate for both independent and dependent experimental variables (i.e. categorical and continuous) and data. (*Referring to the data form) **NI-** If data displayed in a graph is a) not suitable for either the dependent or independent experimental variables OR b) there is a better way to present data. **U-** If the graph type is not suitable for both experimental variables.	x		
	Data Displayed (Raw, Averages, Changes, Percentage) **E-** If the graph indicates the type of data (ex. Raw, averages, etc.) that are plotted. There should be a clear distinction between raw data and manipulated data based on the information presented in the key (ie. sample size and number of trials) and axis label. If the graph is showing averages, then it should also be accompanied with STDEV or error bars. **NI-** If the graph is missing one of points mentioned above. **U-** If data type is inappropriate for the graph type		Since the graph is showing averages, the data points need to be accompanied with error bars.	
	Alignment *(at least one of the graphs presented should align with the research question and hypothesis. Other graphs can be exploratory.)* **E-** If the graph is completely aligned with the research question and/or hypothesis. In other words, the independent, dependent variables, and information about the experiment are explicit. **NI-** If the graph is partially aligned with the research question and/or hypothesis. In other words, the graph is missing information about either the independent, dependent, or details about the experiment. **U-** If the graph is not aligned with the research question and/or hypothesis.		The graph is partially aligned with the research question and hypothesis because it does not display information on the subjects.	

Appendix 4H: Sample Graph 2

Research Question: How does plant growth relate to water intake?
Hypothesis: The more water that is used, the larger the plant will grow.

	ID of graph or set of graphs:	Present/ Appropriate	Present but Needs Improvement	Absent
Graph Mechanics	**Descriptive title** **P/A-** Should be: a) in the form of a statement, b) mention the subject, c) appropriate variables, and d) include relevant details about the experiment that help understand the take home message. **NI-** If the title is missing any one of the four points mentioned above.			x
	Label for the X axis (e.g. time) **P/A-** Should be appropriate and descriptive for the experiment. For graphs with categorical independent variables, there needs to be a label under each set of data and a larger label under all data plotted. **NI-** If the label is missing any one of the points mentioned above.		missing a large x label under the two plots.	
	Label for the Y axis (e.g. heart rate) **P/A-** Should be appropriate and descriptive for the experiment. If the data is manipulated (average, change, percentage, etc.), then it should be indicated on the y axis. **NI-** If the label is missing any one of the points mentioned above.		the y-axis title is not descriptive.	
	Units for the X axis (e.g. seconds) **P/A-** Should be appropriate and descriptive for the data displayed. **NI-** If the units are not appropriate or descriptive.	x		
	Units for the Y axis (e.g. average beats per minute) **P/A-** Should be appropriate and descriptive for the data displayed. **NI-** If the units are not appropriate or descriptive.			x
	Scale (appropriate intervals and range for data) **P/A-** Should be appropriate for the data displayed such that the increments are clear and without clutter and includes appropriate significant figures. If the scale is discontinuous or doesn't start at the origin, it should be indicated by a break in the axis. **NI-** If the scale is not appropriate for the data such that it is cluttered, does not include appropriate significant figures, and/or if the scale does not indicate axis break.	x		
	Key (defines different data sets that are plotted) **P/A-** Should be appropriate and descriptive for the data displayed. It should include: a) descriptions of different colors (if applicable), b) the sample size and c) the number of trials. **NI-** If the key is not descriptive and does not indicate the sample size.		missing sample size for each time point.	



Appendix 4H continued

		Excellent (2)	Needs Improvement (1)	Unsatisfactory (0)
Communication	**Ease of Understanding-Aesthetics** **E**-If the graph is aesthetically pleasing, meaning that: a) the data plotted takes up sufficient room in the Cartesian plane, b) makes use of legible size font, c) the x and y axis lines are clear and legible, d) the graph displays data in an appropriate number of bars and lines, and e) is devoid of chart junk elements such as: distracting background colors, patterns, and dark gridlines **NI**- If the graph has one of the following flaws: a) the graph displays too much white space, b) the font size is too small, c) the x and y axis lines are not clear and legible, d) the graph shows too many bars or lines OR e) elements of chart junk are clouding interpretation of data. **U**- If the graph has multiple flaws, which interfere with the understanding and interpretation of data.	x		
	Ease of Understanding-Take home message **E**- If the graph has sound construction and mechanics that allow for clear sorting of trends and take home message. **NI**- If data trends are difficult to observe or it is difficult to formulate a proper take home message. **U**- If the graph is ineffective at communicating data trends and take home message, such that it causes confusion.		While the graph shows clear trends, it is difficult to formulate a take home message since it is missing a title and various experimental components.	
Graph Choice	**Graph Type (Bar, line, scatter, dot, box and whisker)** **E**- If data displayed in a graph is appropriate for both independent and dependent experimental variables (i.e. categorical and continuous) and data. (*Referring to the data form) **NI**- If data displayed in a graph is a) not suitable for either the dependent or independent experimental variables OR b) there is a better way to present data. **U**- If the graph type is not suitable for both experimental variables.	x		
	Data Displayed (Raw, Averages, Changes, Percentage) **E**- If the graph indicates the type of data (ex. Raw, averages, etc.) that are plotted. There should be a clear distinction between raw data and manipulated data based on the information presented in the key (ie. sample size and number of trials) and axis label. If the graph is showing averages, then it should also be accompanied with STDEV or error bars. **NI**- If the graph is missing one of points mentioned above. **U**- If data type is inappropriate for the graph type	x		
	Alignment *(at least one of the graphs presented should align with the research question and hypothesis. Other graphs can be exploratory.)* **E**- If the graph is completely aligned with the research question and/or hypothesis. In other words, the independent, dependent variables, and information about the experiment are explicit. **NI**- If the graph is partially aligned with the research question and/or hypothesis. In other words, the graph is missing information about either the independent, dependent, or details about the experiment. **U**- If the graph is not aligned with the research question and/or hypothesis.			The graph does not align with the RQ or HYP. Since the details on the experiment(specifically the water amounts) are not provided, and the axes are not labelled properly, it is difficult to link day 0 and day 10 to a particular water treatment.

Appendix 4H: Sample Graph 3

Researcher's Name	AA	
Book Citation		ICB
Chapter		9
Page Number		
Figure #		9.22 (top)

	ID of graph or set of graphs:	**Present/ Appropriate**	**Present but Needs Improvement**	**Absent**
Graph Mechanics	**Descriptive title** P/A-Should be: a) in the form of a statement, b)	fig legend has all the info.		
	Label for the X axis (e.g. time) P/A- Should be appropriate and descriptive for	x		
	Label for the Y axis (e.g. heart rate) P/A- Should be appropriate and descriptive for		avg.	
	Units for the X axis (e.g. seconds) P/A- Should be appropriate and descriptive for	x		
	Units for the Y axis (e.g. average beats per minute) P/A- Should be appropriate and	x		
	Scale (appropriate intervals and range for data) P/A- Should be appropriate for the data	x		
	Key (defines different data sets that are plotted) P/A- Should be appropriate and descriptive for		sample size?	
		Excellent (2)	**Needs Improvement (1)**	**Unsatisfactory (0)**
Communication	**Ease of Understanding-Aesthetics** E-If the graph is aesthetically pleasing, meaning NI- If the graph has one of the following flaws: a U- If the graph has multiple flaws, which interfer	x		
	Ease of Understanding-Take home message E- If the graph has sound construction and mech NI- If data trends are difficult to observe or it is U- If the graph is ineffective at communicating d	x		
Graph Choice	**Graph Type (Bar, line, scatter, dot, box and w** E- If data displayed in a graph is appropriate for NI- If data displayed in a graph is a) not suitable U- If the graph type is not suitable for both exper	x		
	Data Displayed (Raw, Averages, Changes, Pe E- If the graph indicates the type of data (ex. Ra NI- If the graph is missing one of points mention U- If data type is inappropriate for the graph typ	avg with SD		
	Alignment *(at least one of the graphs present* E- If the graph is completely aligned with the re NI- If the graph is partially aligned with the rese U- If the graph is not aligned with the research q			

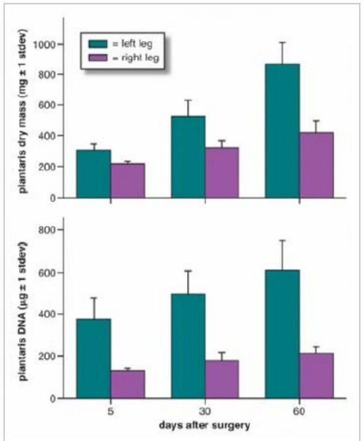


Figure 9.22 Effect of removal of the gastrocnemius muscle on the plantaris muscle in rats. **A,** Plantaris dry mass + SD. **B,** Total muscle DNA + SD. Data from Hubbard *et al.*, 1975. Table 1; original art.

Appendix 4H: Sample Graph 4

Researcher's Name
Book Citation
Chapter 58
Page Number 1217
Figure # 58.8

	ID of graph or set of graphs:	Present/ Appropriate	Present but Needs Improvem	Absent
Graph Mechanics	**Descriptive title** P/A-Should be: a) in the form of a statement, b) mention the subject, c) appropriate variables, and d) include relevant details about the experiment that help understand the take home message. NI- If the title is missing any one of the four points mentioned above.	fig legend contains all the details.		
	Label for the X axis (e.g. time) P/A- Should be appropriate and descriptive for the experiment. For graphs with categorical independent	x		
	Label for the Y axis (e.g. heart rate) P/A- Should be appropriate and descriptive for the experiment. If the data is accumulated (average	x		
	Units for the X axis (e.g. seconds) P/A- Should be appropriate and descriptive for the data displayed	x		
	Units for the Y axis (e.g. average beats per minute) P/A- Should be appropriate and descriptive for the data displayed	x		
	Scale (appropriate intervals and range for data) P/A- Should be appropriate for the data displayed such that the increments are clear and without clutter	x		
	Key (defines different data sets that are plotted) P/A- Should be appropriate and descriptive for the data displayed. It should include: a) descriptions of	x		
		Excellent (2)	**Needs Improvement (1)**	**Unsatisfactory (0)**
	Ease of Understanding-Aesthetics E-If the graph is aesthetically pleasing, meaning that: a) the data plotted takes up sufficient room in the Cartesian plane, b) makes use of legible size font, c) the x and y axis lines are clear and legible, d) the graph displays data in an appropriate number of bars and lines, and e) is devoid of chart junk elements such as: distracting background colors, patterns, and dark gridlines	x		

Appendix 4H continued

Communication	**NI**- If the graph has one of the following flaws: a) the graph displays too much white space, b) the font size is too small, c) the x and y axis lines are not clear and legible, d) the graph shows too many bars or lines OR e) elements of chart junk are clouding interpretation of data. **U**- If the graph has multiple flaws, which interfere with the understanding and interpretation of data.			
	Ease of Understanding-Take home message **E**- If the graph has sound construction and mechanics that allow for clear sorting of trends and take home message. **NI**- If data trends are difficult to observe or it is difficult to formulate a proper take home message. **U**- If the graph is ineffective at communicating data trends and take home message, such that it causes confusion.	x		
Graph Choice	**Graph Type (Bar, line, scatter, dot, box and whisker)** **E**- If data displayed in a graph is appropriate for both independent and dependent experimental variables (i.e. categorical and continuous) and data. (*Referring to the data form) **NI**- If data displayed in a graph is a) not suitable for either the dependent or independent experimental variables OR b) there is a better way to present data. **U**- If the graph type is not suitable for both experimental variables.	x		
	Data Displayed (Raw, Averages, Changes, Percentage) **E**- If the graph indicates the type of data (ex. Raw, averages, etc.) that are plotted. There should be a clear distinction between raw data and manipulated data based on the information presented in the key (ie. sample size and number of trials) and axis label. If the graph is showing averages, then it should also be accompanied with STDEV or error bars. **NI**- If the graph is missing one of points mentioned above.	x		

	U- If data type is inappropriate for the graph type			
	Alignment *(at least one of the graphs presented should align with the research question and hypothesis. Other graphs can be exploratory.)* **E**- If the graph is completely aligned with the research question and/or hypothesis. In other words, the independent, dependent variables, and information about the experiment are explicit. **NI**- If the graph is partially aligned with the research question and/or hypothesis. In other words, the graph is missing information about either the independent, dependent, or details about the experiment. **U**- If the graph is not aligned with the research question and/or hypothesis.			

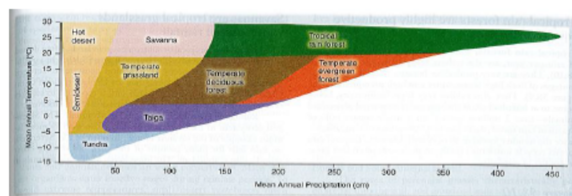


figure 58.8
PREDICTORS OF BIOME DISTRIBUTION. Temperature and precipitation are quite useful predictors of biome distribution, although other factors sometimes also play critical roles.

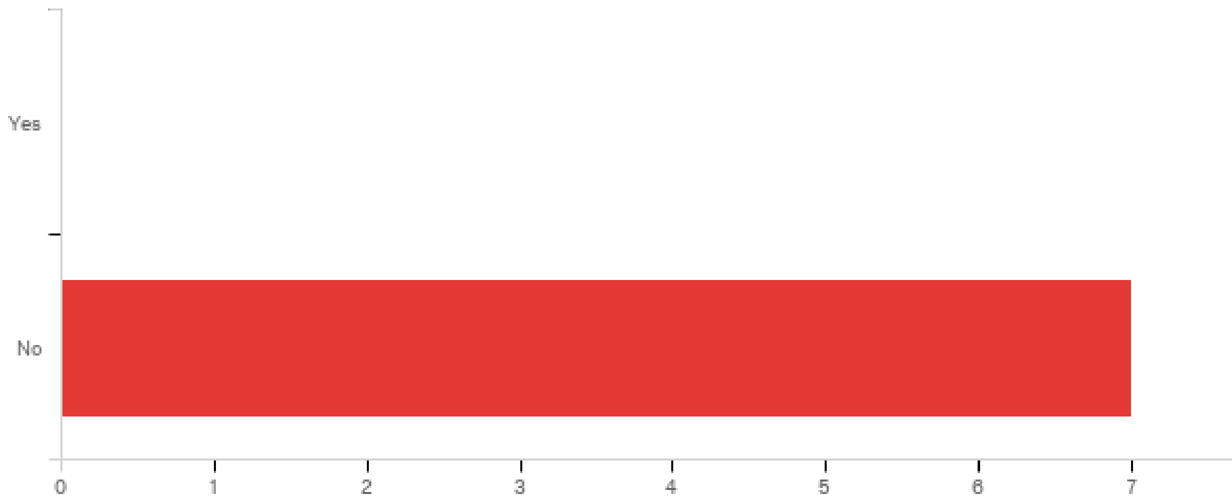
Appendix 4I: Feedback from Rubric Validation from 3 Professors

Initial Report

Evaluating Graphs with Rubric-Physiology Students

August 20th 2016, 3:06 pm EDT

Q2 - Did you encounter any problems with the rubric? Please explain.



Answer	%	Count
Yes	0.00%	0
No	100.00%	7
Total	100%	7

Yes

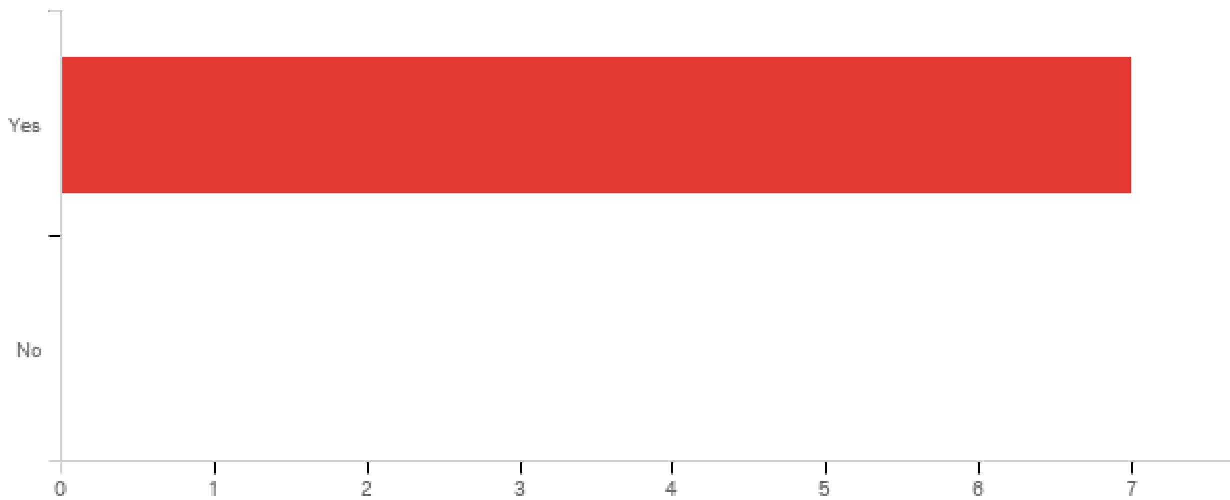
Yes

Appendix 4I continued

No

No
Very clear
I thought this was more clear then the other graph
no problems were encountered

Q3 - Was the language in the rubric easy to understand? Please explain.



Answer	%	Count
Yes	100.00%	7
No	0.00%	0
Total	100%	7

Yes

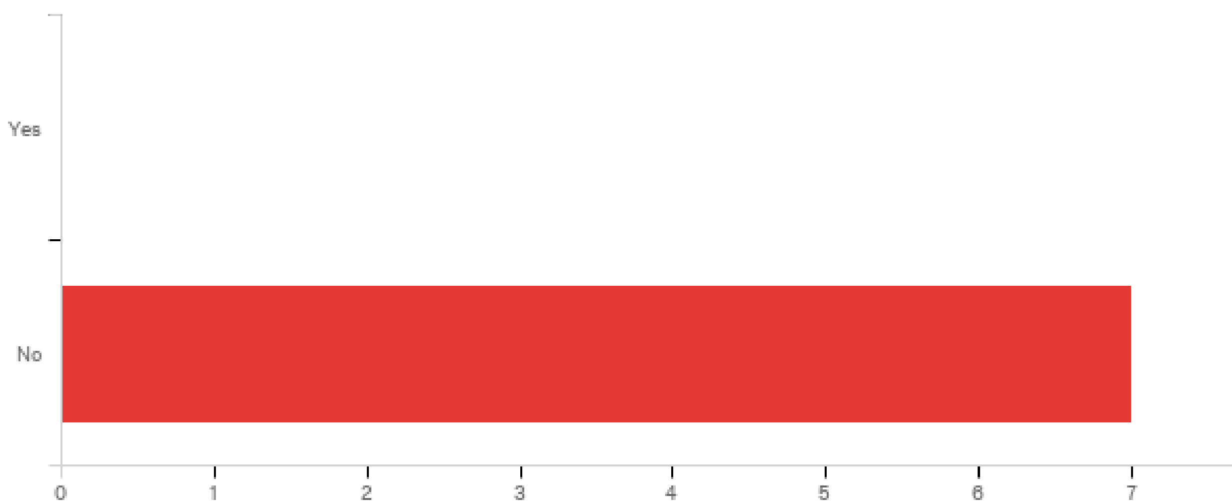
Appendix 4I continued

Yes
Clear instructions
The specific points detailing what should be present in the graph were helpful.
very straight forward language
Everything was very clear t

No

No

Q4 - Do you have suggestions for improvements? Please explain.



Answer	%	Count
Yes	0.00%	0
No	100.00%	7
Total	100%	7

Appendix 4I continued

Yes

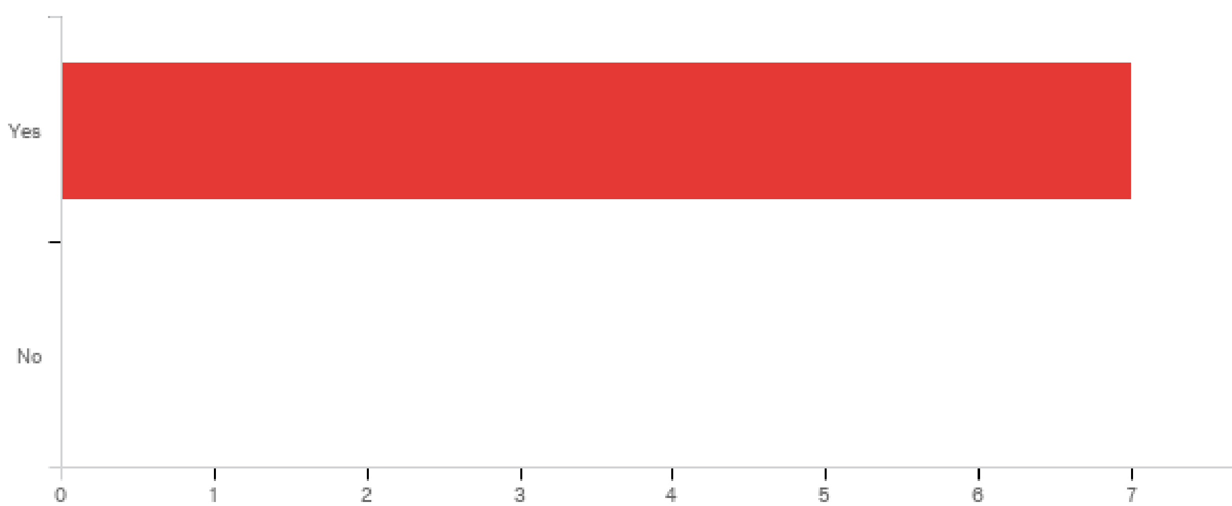
Yes

No

No

I thought it was clear and easy to understand
I like this one a lot
seems to cover the necessary points
Have different amount of points taken off depending on what is missing. The needs improvement column is -1 but if there was only a small mistake it should be less

Q5 - Is this rubric useful for evaluating graphs? Please explain.



Answer	%	Count
Yes	100.00%	7

Appendix 4I continued

No	0.00%	0
Total	100%	7

Yes

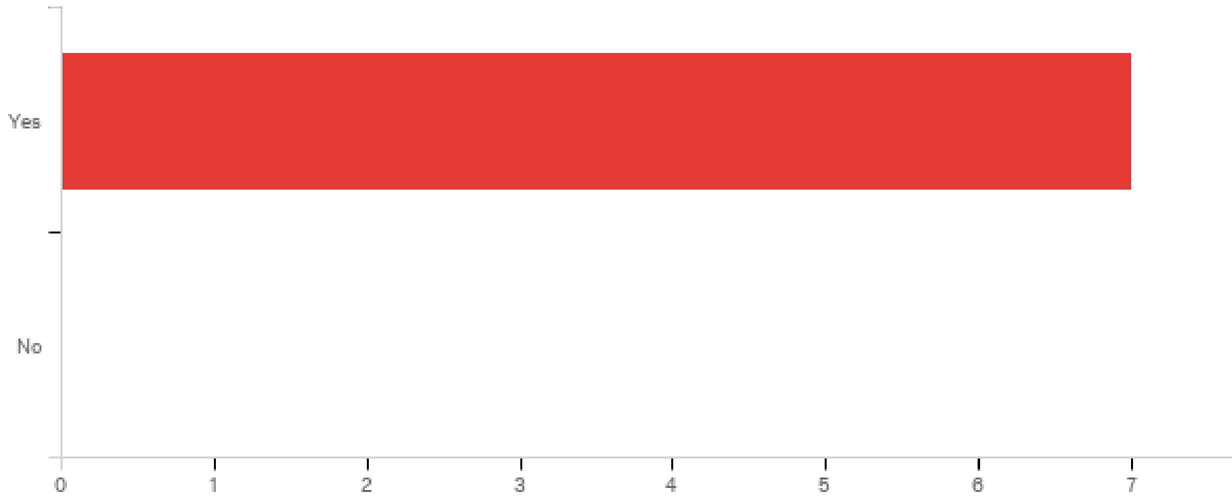
Yes
Helps to notice small things that can be improved that you may not notice without the rubric. Also helps you break down and locate the problems in particularly messy graphs
Good examples of grading comments
The rubric is helpful to serve as a reminder for everything that needs to be included in a graph.
focused on the necessary elements needed for understanding a graph without excessive explanations
Yes. It really breaks down what is needed in the graphs

No

No

Appendix 4I continued

Q6 - Would you use this rubric in your other classes? Please explain.



Answer	%	Count
Yes	100.00%	7
No	0.00%	0
Total	100%	7

Yes

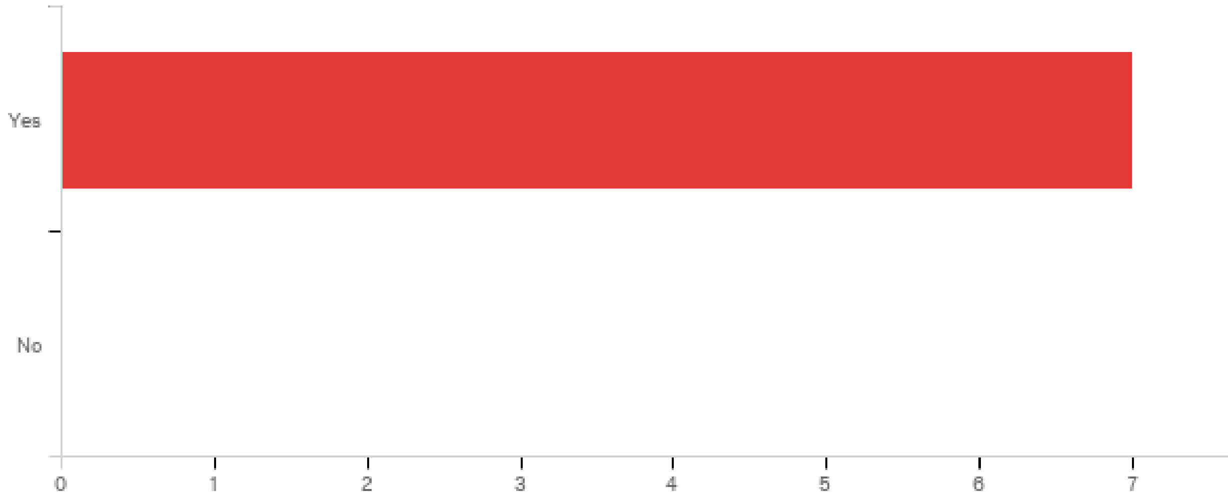
Yes
Good examples
This would be helpful to me to make sure I have everything I need on my graph
This rubric is useful as a guideline for creating effective graphs.
the rubric is both comprehensive and flexible enough to be used in other scientific courses
Yes, especially in my science courses

Appendix 4I continued

No

No

Q8 - Would you recommend this rubric to your friends? Please explain.



Answer	%	Count
Yes	100.00%	7
No	0.00%	0
Total	100%	7

Yes

Yes
I think it can help you go back and review your own graphs
I would reccommend this to friends in science classes who are making graphs similar to ones used in physiology

Appendix 4I continued

I think the rubric could help my peers focus on what is important to include in a graph.
it can show other classmates how to improve their own graphs or evaluate others' graphs (other classmates or in scientific articles)
Yes, before they turn something in and want to make sure they have included everything

No

No

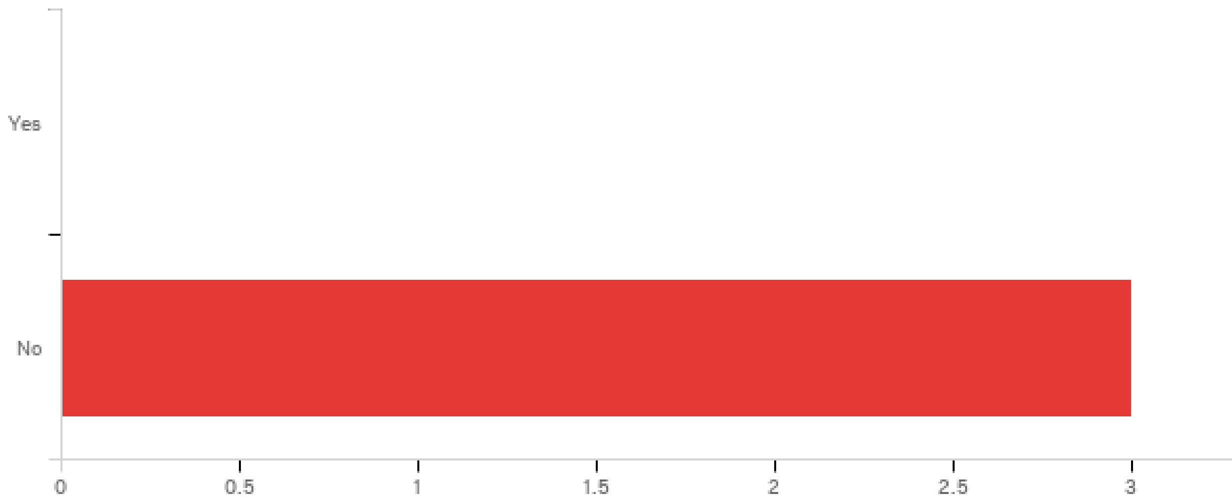
Appendix 4J: Feedback from Rubric Validation from 3 Professors

Initial Report

Evaluating Graphs with Rubric-Experts in Biology

August 20th 2016, 3:13 pm EDT

Q2 - Did you encounter any problems with the rubric?



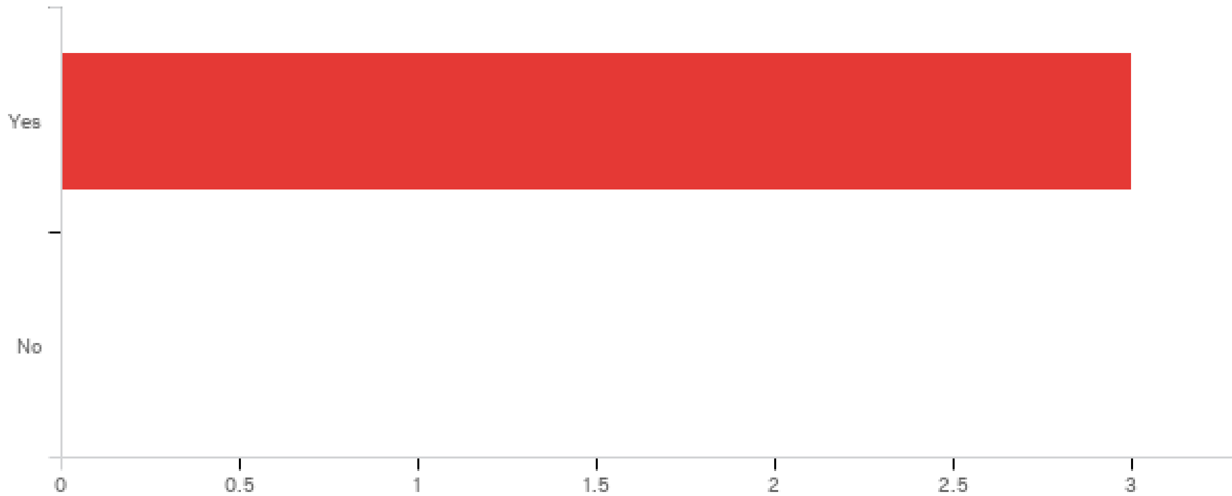
Answer	%	Count
Yes	0.00%	0
No	100.00%	3
Total	100%	3

Yes

Y es

Appendix 4J continued

Q3 - Was the language in the rubric easy to understand?



Answer	%	Count
Yes	100.00%	3
No	0.00%	0
Total	100%	3

Yes

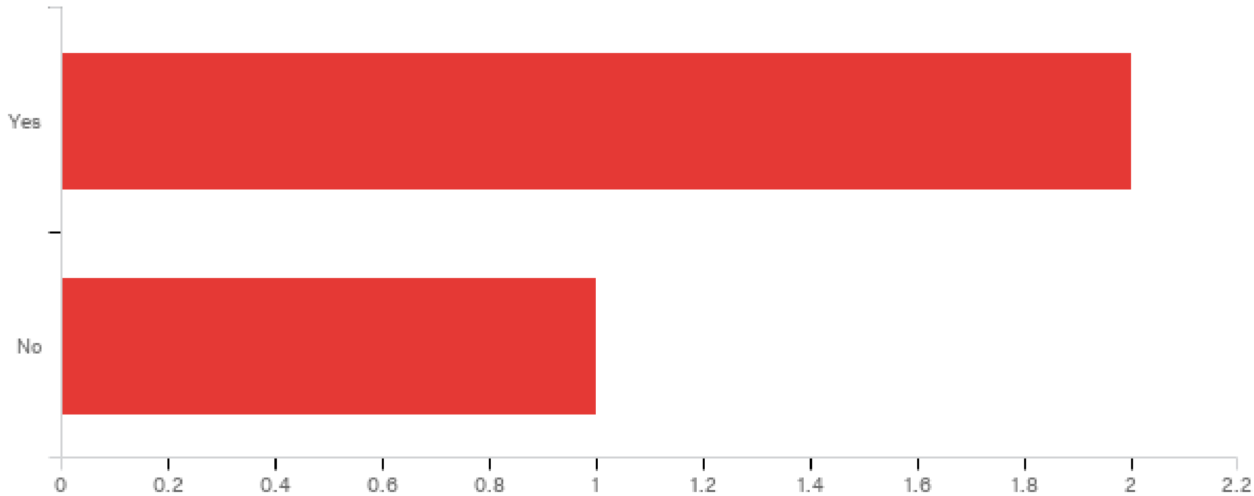
Yes

No

No

Appendix 4J continued

Q4 - Do you have suggestions for improvements?



Answer	%	Count
Yes	66.67%	2
No	33.33%	1
Total	100%	3

Yes

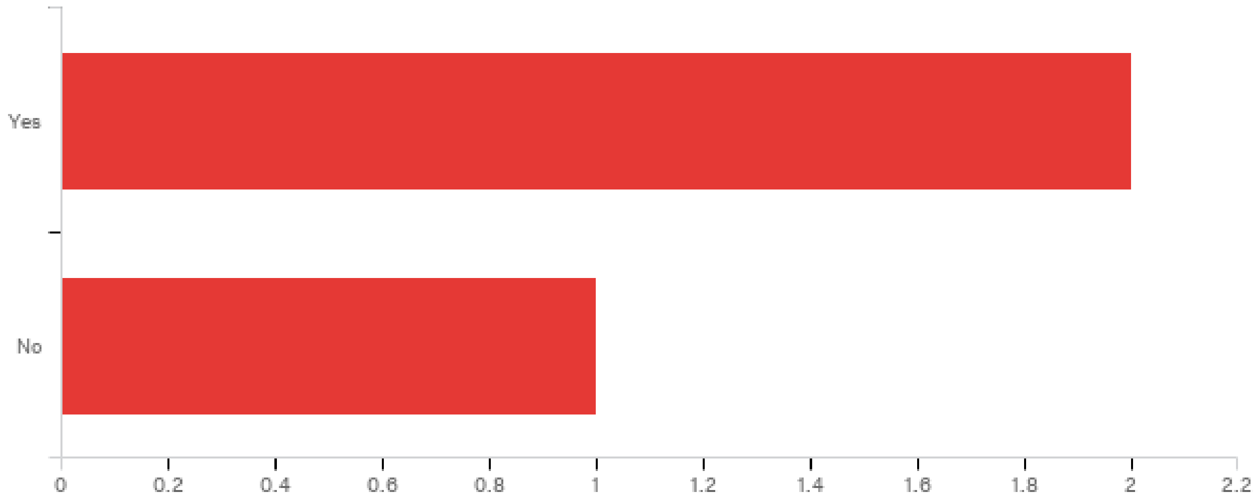
Yes
Give user the chance to define purpose and hypothesis

No

No

Appendix 4J continued

Q5 - Would you implement this rubric in your class(es)?



Answer	%	Count
Yes	66.67%	2
No	33.33%	1
Total	100%	3

Yes

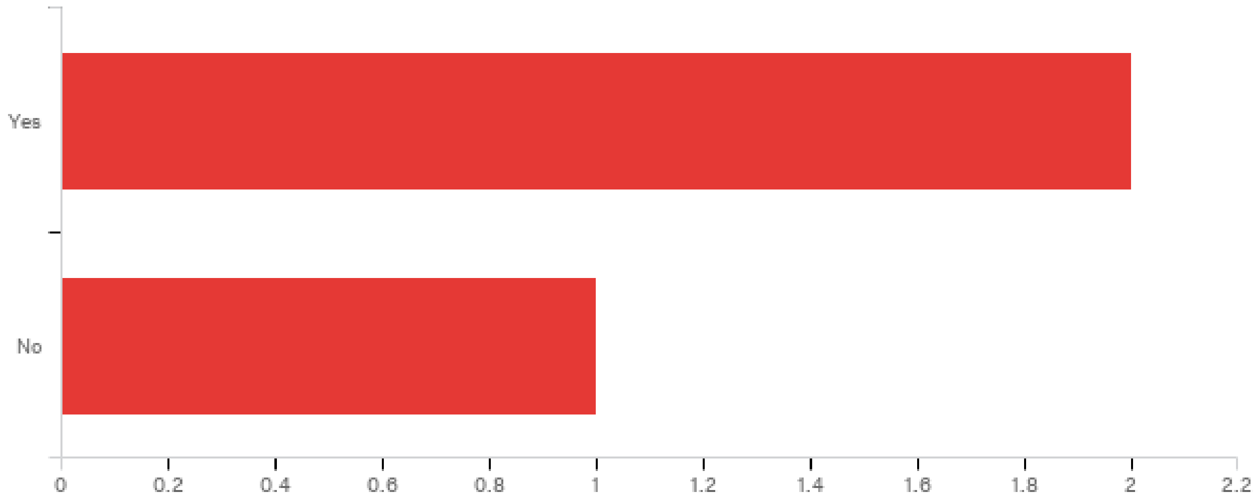
Yes

No

No

Appendix 4J continued

Q6 - Do you think students would find this rubric useful?



Answer	%	Count
Yes	66.67%	2
No	33.33%	1
Total	100%	3

Yes

Yes

No

No

CHAPTER 5: USING THE COGNITIVE APPRENTICESHIP MODEL TO IMPROVE UNDERGRADUATE BIOLOGY STUDENTS' GRAPH CHOICE AND CONSTRUCTION IN AN UPPER-DIVISION PHYSIOLOGY LABORATORY CLASSROOM

5.1 Abstract

The objective of the work described in this chapter was to incorporate evidence-based instructional materials (chapter 4) into a semester-long instructional intervention using the cognitive apprenticeship model (CAM). This model aims to make implicit knowledge explicit, to make the contexts of student learning authentic and relevant, and to provide the learner with a variety of experiences and contexts in which to practice skills, which guides them towards expertise. There are 6 components in CAM: (1) modeling of instructional materials by the instructor, (2) coaching the students through observations and offering suggestions, (3) scaffolding the learning by providing guidance to the students, (4) allowing students to articulate their knowledge and thinking, (5) engaging the students in reflective tasks that allow them to compare and modify their knowledge with that of the instructor and/or other students, and (6) challenging students to explore novel contexts and apply their knowledge. The context of this study was an inquiry-based laboratory component of an upper division physiology course taken by biology majors at a large, Midwestern research-intensive university (N=123 students) during the Spring 2015 and 2016 semesters. In this class students worked in small groups to design experiments, collect and analyze data, and present their findings to the

class. Student learning of graphing was facilitated by applying the previously developed instructional resources (chapter 4) within the CAM components.

Frequency of downloads of instructional materials was tracked from course management software throughout the semester. The effectiveness of the teaching intervention was evaluated by: pre/post survey on graph knowledge, and comparison of students' graphs to those in the non-intervention semesters, Spring 2013 and 2014. The pre and post surveys showed students' greater awareness of various types of graphs in the post survey, but the line graph was a popular choice during both the pre and post surveys. There was no observable change between the pre and post surveys with regards to reasoning behind graph choice. Compared to student groups in the non-intervention semesters, by the last lab, more student groups chose to construct either a box or dot plot, graphs were appropriately aligned with the research question and hypothesis, and the quality of graphs constructed was better in the intervention semesters. In addition, this instructional approach, with its resources and practices, has provided further insights into student competencies and difficulties and can be used to guide future instruction.

5.2 Overview

Quantitative skills are necessary components of the undergraduate biology curricula because they help undergraduate students prepare for medical school (AAMC/HHMI, 2009), graduate school (Barraquand et al., 2014), and are part of general science literacy (Gormally, Brickman, & Lutz, 2012). Constructing and interpreting graphs are some of the quantitative skills recognized and taught by faculty (Gormally, Brickman, & Lutz, 2012). In previous chapters of this dissertation, we revealed the thought processes used by novice and expert biologists when constructing graphs in a think-aloud interview (chapter 2) and reasoning used by students when choosing and constructing graphs of their experimental data collected in a classroom (chapter 3). Based on data from these two chapters, we developed three evidence-based graphing materials for classroom use (chapter 4). The purpose of this chapter is to test the utility of these evidence-based graphing materials for improving undergraduate student competence with graphing in a classroom environment.

5.3 Background

Successful construction and reasoning with external representations, such as graphs, requires students to be knowledgeable and reflective in four areas: invention, critique, functioning, and learning or reflection. These four components comprise meta-representational competence, or MRC (diSessa & Sherin, 2000; diSessa, 2004) which describes the knowledge and skills of a competent representation maker. The first component, invention, explains how individuals are able to conceive novel graphical representations from data (diSessa & Sherin, 2000). The second component, critique, exposes students' critical knowledge for assessing strengths and weaknesses of various types of graphs (diSessa & Sherin, 2000). The third component, functioning, explains individuals' underlying reasoning for connecting the function and usage of various graphs on the type of data present (diSessa & Sherin, 2000). The final component, learning or reflection, is a strategy that fosters individuals' awareness of their own understanding of graphs and gaps in knowledge (diSessa & Sherin, 2000).

Self-reflection and critical thinking are some of the best ways for students to evaluate and think about their learning (Tynjala, 1999). Although oral reflections via interviews are a rich source of data to convey to the researcher what the student is thinking, they are oftentimes not practical in a naturalistic setting. Written reflections are useful when they probe students to think deeply. In order to increase students' confidence and refine their critical thinking skills, and the learning or reflection component of the MRC, reflections should be performed and encouraged numerous times throughout a class.

Findings presented in chapters 2 and 3 of this dissertation have revealed areas of undergraduate biology student competencies and difficulties with graphing. As previously reported by Padilla, McKenzie and Shaw (1986), undergraduate biology students in our studies also showed competency with plotting data points and determining the X and Y coordinates of data (chapter 2). We also saw previously documented difficulties at the K-12 levels with undergraduate biology students, such as graph choice (Leonard & Patterson, 2004; Tairab & Al Naqbi, 2010) labelling title and axes (Leonard & Patterson, 2004), understanding variables (Tairab & Al Naqbi, 2010), scaling axes (Padilla et al., 1986; Brasell & Rowe, 1993; Ainley, 1995), representing raw data accurately (Mevarech & Kramarsky, 1997; Bakker, 2004; Leonard & Patterson, 2004; Meletiou & Lee, 2010; Bray-Speth et al., 2010), understanding and plotting slope and y-intercept (Hattikudur et al., 2012) , performing simple calculations (Meletiou & Lee, 2010; Bray-Speth et al., 2010). One reason for these continued difficulties into the undergraduate education could be due to the paucity of research around elements of successful teaching interventions with undergraduate graphing. In the context of undergraduate biology, there have been two studies that provide insight into student difficulties with data and graphing (Bray-Speth et al., 2010; McFarland, 2010) McFarland (2010) designed a laboratory class intervention devoted to graph choice and construction that included instructions on graphs, their usage, and student engagement in the reflection on appropriateness of graphs. Over the semester, students were asked to reflect on their graph choice and quality, engaging the learning or reflection component of MRC. Although students were not explicitly assessed on their graphs or reasoning, students' graphing quality improved over time and students responded positively to a course evaluation question about their learning about graphs in the class. Unlike the McFarland (2010) study, the Bray- Speth et al. (2010) study occurred in the context of an introductory undergraduate biology lecture class. Students showed significant improvements in graph mechanics, but there was an inconsistency in graph choice. Although these studies report interesting and useful findings in the undergraduate biology context, it is challenging to deduce best practices that promote the development of graphing proficiency. Our study aims to understand the implementation of evidence-based graphing materials developed in chapter 4 of this dissertation and any additional instructional approaches that could be used to target and improve previously documented graphing difficulties in undergraduate students.

5.4 Instructional Model and Theoretical Framework

This research is guided by the cognitive apprenticeship model (CAM), which is a social learning method that helps novices transition towards expertise (Dennen, 2004). The idea of apprenticeship, where the novice acquires knowledge, concepts, and skills to master a trade or craft, has been practiced throughout history, even before this model was adapted to formal education in 1989 by Collins et al. (1991) (Dennen, 2004). Unlike the earlier apprenticeship models which were used for trade purposes that stressed the importance of physical skills, the CAM focuses on developing cognitive and metacognitive skills by participating in authentic learning experiences (Collins et al., 1991; Dennen, 2004). The CAM consists of 6 components that draw on the social constructivist learning theory, developed by Lev Vygostsky (1962), which emphasizes the collaborative nature of learning, reiterating interaction between the instructor and learner: (1) modeling of the instructional materials and processes of thinking by the instructor, (2) coaching the students through observations and offering suggestions, where necessary (3) scaffolding the learning by providing guidance to the students, (4) allowing students to articulate their knowledge and thinking, (5) engaging the students in reflective tasks that allow them to compare and modify their knowledge with that of the instructor and/or other students, and (6) encouraging students to formulate and test new research questions and hypotheses, in order to apply their knowledge. Steps in the CAM model such as articulation, reflection, and exploration are habitual to experts and helps students achieve mastery (Hogan & Maglienti, 2001). A summary of the six steps, along with the pros and cons of each step are displayed in Table 5.1 (Collins et al., 1991; Collins, 1996).

The context in which the CAM is applied is vital to its success in transferring knowledge from experts to novices. Active student participation and repeated practice is ideal since the instructional model is based on the premise that involvement in the subject matter in an authentic situated learning environment encourages the novice to learn and aids in accelerating them towards expertise (Dennen, 2004; McFarland 2010; Roth & Bowen 2001).

The CAM is an appropriate model of instruction for our upper-division inquiry-based lab, because inquiry labs work to emulate habits of practicing scientists (Hogan & Maglienti, 2001). In this lab, students were actively engaged in the process of designing experiments, collecting data, and constructing graphs. In addition to pursuing their roles as instructors, the nature of this lab allowed the instructors to serve as coaches and mentors to the students. Instructors in this setting usually engage in conversations with students, coach students, and have them articulate their thinking, in order to better guide them towards expertise (Gormally et al., 2009).

Table 5.1: Steps in the Cognitive Apprenticeship Model (CAM)

Steps in CAM	Definition	Implementation in our Study	Pros	Cons
Modeling	Modeling the thought processes that underlie expert performance. (Collins et al., 1991; Collins, 1996)	In week 2, we introduced students to the instructional materials, told them the importance of the materials, where they can be accessed throughout the semester, our expectations for how we would like students to use them, and gave students practice using these materials. After students had a chance to practice with these materials, we engaged them in a discussion, during which time we revealed our underlying thought processes.	Usually invisible, students can integrate bits and pieces and understand why it happens. (Collins, 1996)	This activity is more instructor focused, so for students, it is a passive activity that is boring. (Collins, 1996)
Scaffolding	Support given to the students as they carry out a task. This is much like short skis that enable people to ski much faster or cue cards that help students write. (Collins et al., 1991; Collins, 1996)	We designed the step-by-step guide and graph rubric to guide students towards appropriate graph choice and construction. These materials focused students to the key elements that we wanted them to learn.	Helps students accomplish difficult tasks providing focused help at critical times. (Collins, 1996)	Crutch that students know they can fall back on, so they may become dependent on it. (Collins, 1996). So with regards to the graphing materials, students may be hesitant to automatize their thinking if they do not have these graphing materials with them.
Coaching	Involves multiple activities like choosing tasks, modeling how to do them, providing hints and scaffolding, diagnosing problem, giving feedback, challenging and offering encouragement and structuring the way to do things. Ex. Computer programs that see how a student is doing a certain task, and then gives feedback on improving the score. (Collins et al., 1991; Collins, 1996)	Throughout the semester, we encouraged students to use the step-by-step guide, guide to data displays, and the graph rubric to help them when thinking about their data and the types of graphs to make. We also helped students with graphing software such as Excel and Origin when making their graphs. Also, throughout the semester, we gave students feedback on their graphs with the graph rubric.	Give students focused help at critical times. (Collins et al., 1991; Collins, 1996)	Misdiagnosis especially giving students feedback on the computer, because you can't see student behavior. (Collins, 1996)
Articulation	Having students articulate ideas and thinking processes. (Collins et al., 1991; Collins, 1996)	Throughout the semester, the instructors circulated around the classroom to encourage students to explicitly articulate their reasoning for graphing data. Instructors also involved other group members in the dialogue to encourage collaborative thinking. Students also articulated their ideas when presenting their PowerPoint to the class.	Helps make people's tacit knowledge explicit so that it is more available. Allows other people to see how other people think about the same problem making knowledge more available through articulation fosters transfer of this knowledge to new situations. (Collins, 1996)	Students may learn to memorize specifics about the graphing material without really understanding the purpose of graphs. This may also discriminate amongst the quiet students who can do tasks perfectly without any articulation. (Collins, 1996)
Reflection	Involves looking back over one's performance on a task and comparing it to other people's performances, both good and bad on similar tasks. (Collins et al., 1991; Collins, 1996)	Four times over the course of the Spring 2015 and 2016 semesters, students utilized the graph rubric to critique each other's graphs. The reason for this exercise was twofold. First, we wanted to expose students to graphs created by their peers and allow them to practice all four components of MRC while engaging in this exercise. The second reason was by engaging in this reflective exercise, would allow students to be aware of their own thought processes and understanding so that when they are constructing graphs for future labs, they can recall the graphs they reflected on, with a better understanding about the advantages and disadvantages of graphs. We hoped that this exercise would improve students' overall graph choice.	Students have the chance to see the process for the first time and compare themselves to others; they can see themselves from a new angle. (Collins, 1996)	Students may find it boring to look back through graphs their peers just presented and may lack patience to thoughtfully reflect. Students may not see the value in this activity and just want to move on to other tasks. (Collins, 1996)
Exploration	Aimed at encouraging independence at executing expert problem-solving processes. (Collins et al., 1991)	Four times over the course of the Spring 2015 and 2016 semesters, students were challenged with formulating research questions, hypotheses, designing experiments, collecting data, and presenting their findings in graphs. We wanted students to use the knowledge they had acquired through the graphing materials, and to use peer and instructor feedback on their graphs to improve their graphing approaches as they practiced these skills in new contexts throughout the semesters.	Allows students to become independent thinkers by having opportunities to practice expert-like thinking and applying it to new contexts (Collins et al., 1991).	Instructors need to teach exploration strategies as part of the learning strategies to successfully transition students away from scaffolds and towards independent thinking (Collins et al., 1991). Students need to be motivated to apply their knowledge to new contexts.

5.5 Research Questions

The overarching objective of this study is to implement the evidence- based graphing materials, step-by-step guide, guide to data displays, and the graph rubric in a classroom setting to see if it has a positive impact on students' graph reasoning and construction. In the process of implementing the graphing materials, we also validated them with the students enrolled in the Spring 2015 and 2016 semesters. To accomplish this objective, we sought to answer the following questions:

1. Does the application of the Cognitive-Apprenticeship Model, (CAM) in an upper-level undergraduate physiology laboratory with associated instructional materials and approaches, improve students' graph choice and construction over the intervention semesters as compared to non-intervention semesters?

   1a) Does consistent use and targeted feedback with the graph rubric improve graph mechanics, communication, choice, and alignment with the research question and hypothesis during the non-intervention and intervention semesters?

   1b) Does usage of the guide to data displays improve students' awareness of various graph types over the course of the semester?

5.6 Methods

The setup of the labs, methods for group assignment, instructors, and points associated with assignments and presentations in Spring 2015 and 2016 semesters were the same as the preceding Spring 2013 and 2014 semesters because we wanted to use the semesters where evidence-based graphing materials were not implemented as a comparison group. In order to implement the graphing materials under the CAM, we made three changes. The first change was administering the evidence-based graphing materials in the inquiry-based laboratory. Although students in Spring 2013 and 2014 had access to a graph rubric (See Appendix 3C) and an earlier version of the guide to data displays (See Appendix 3B), these materials were neither evidence based nor detailed. The second change was using the CAM in the inquiry-based lab to inform the instruction. In the previous Spring 2013 and 2014 semesters, the role of the instructor was still that of a facilitator, but the CAM provided structure for the instructors by allowing them to model the evidence-based graphing materials and use these materials as scaffolds when coaching students. The third change was students in Spring 2015 and 2016 did not reflect on their own graphs, but on graphs constructed by other groups. We explain this idea below in the methods. We also revisited the data reported in chapter 3 and provide a deeper evaluation on the types of graphs constructed, using the graph rubric as an assessment tool.

Table 5.2: Demographics of the students enrolled in the upper-level physiology laboratory during the non-intervention (Spring 2013 and 2014) and intervention (Spring 2015 and 2016) semesters.

Semester	Year	# of Students Enrolled	Gender (%)		Undergraduate Major (%)			Hours Class Standing (%)					
			Male	**Female**	**General Biology**	**Neurobiology**	**Other**	**0-14 hours**	**45-59 hours**	**60-74 hours**	**75-89 hours**	**90-104 hours**	**105+ hours**
Non-Intervention	**2013**	72	41.67	58.33	47.22	34.72	18.06	0	0	0	2.78	9.72	87.50
	2014	67	43.28	56.72	19.40	53.73	26.87	0	0	0	14.90	23.39	62.69
Intervention	**2015**	66	28.79	71.21	36.36	37.88	25.76	3.03	0	0	16.67	28.79	51.52
	2016	57	35.09	64.91	33.33	42.11	24.56	0	1.75	1.75	15.79	7.02	73.68

5.6.1 Participants

This study was conducted in an upper-level principles in physiology laboratory at a large midwestern university during the spring 2015 and 2016 semesters. Table 5.2 shows the demographics of the students enrolled in the upper-level physiology laboratory during the Spring 2013, 2014, 2015 and 2016 semesters. From hereafter, we refer to the students and data reported from Spring 2013 and 2014 as the "non-intervention semester" and the students and data collected from Spring 2015 and 2016 as the "intervention semester". Since one of the goals of the study described in this chapter is to evaluate the impact of graphing materials, we compare students from the non-intervention semester (these students did not receive the validated graphing materials) to students in the intervention semester, who had full exposure and unlimited usage of these materials. The layout of the lab activities are shown in Figure 5.1 with notations on components of the CAM. All laboratory experiments were inquiry-based, open-ended, and allowed students to work together in teams of four to design experiments, collect data, and present data in a 10 minute PowerPoint presentation. Similar to the methodology in chapter 3, this process occurred five times over the course of each of the 2015 and 2016 semesters, but we report data from the four labs in which students had adequate time to complete the reflections or peer evaluations (see below).

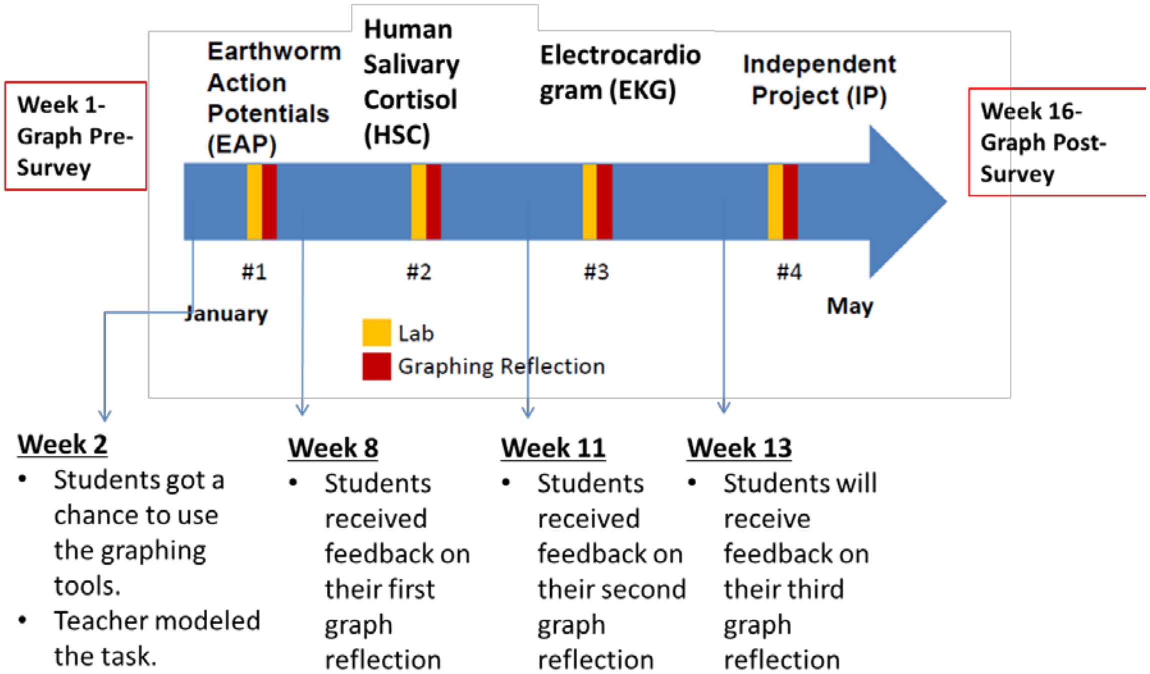


Figure 5.1: Implementing CAM in Spring 2015 and 2016 Semesters: The timeline above illustrates how the CAM was utilized in an upper-level biology classroom. Students took a pre-survey on graphing the first week of lab. Formal instruction on the graphing materials began in the second week of the semester, with modeling of the graphing materials and having students practice using these materials. Throughout the semesters, instructors coached, provided scaffolds, and encouraged exploration by inviting students to pose and solve their own problems. After students articulated their experimental findings orally via PowerPoint presentations, they were asked to use the graph rubric to critique peer graphs and reflect on their own knowledge of the advantages and disadvantages of graphs.

Under an exempt IRB protocol (#1208012562, see Appendix 3H), we report on graphs constructed by student groups from four labs (Figure 5.1) over the course of the intervention semesters. We also report qualitative data from pre and post surveys from 60 undergraduate students from Spring 2015 and 49 undergraduate students from Spring 2016. At the time of study, the majority of the students enrolled in the 2015 and 2016 semesters were either pursuing bachelors in biology in neurobiology and physiology or general biology (See Table 5.2). Although the majority of the students enrolled in the course were rising juniors or seniors, as previously established in Chapter 3, we assumed all physiology students were at the journeyman level when they signed up for this course (Bing & Redish, 2012). This classifies the students midway between the expert and novice level, because they have the fundamental knowledge and skills that does not make them novices, but their lack of practice with abstruse problem solving also does not categorize them with the experts (Bing & Redish, 2012). Hence, we report data from students at the journeyman level, who utilized the three graphing materials over the intervention semesters.

5.6.2 Laboratory Context

As with the previous non-intervention semesters, the intervention semester labs are inquiry-based in design. Each lab began with a short PowerPoint reinforcing the concepts learned in lecture prior to lab and providing a brief overview of the lab objectives, equipment used, as well as the expectations from students. The same tool developed by Ohland and colleagues (2006), called CATME (Comprehensive

Assessment of Team Member Effectiveness) was used to assign students into groups for the intervention semesters. Our expectations from students and groups did not change for the intervention semester. Students were expected contribute equally towards the team to formulate research questions, hypotheses, design experiments, collect data, and present findings in a short 10 minute PowerPoint presentation to the class. All lab activities were open-inquiry and project-oriented labs that required students' to actively engage and practice science process skills such as: formulating a research question and hypothesis, planning and designing an experiment for that weeks' lab topic, identifying and defining key variables, transforming, analyzing, and interpreting data (Padilla, 1990; Roth & Roychoudhury, 1993). Practicing science process skills in this open context leads to students answering questions that they are naturally curious about, along with becoming independent and responsible learners, which leads to student competency with these process skills (Roth & Roychoudhury, 1993). This lab context has proved beneficial for not only the strong students, but also the weaker students (Roth & Roychoudhury, 1993). Although we are not analyzing individual students or evaluating their science process skills, we are focusing on a small but important piece of these process skills, which involves data handling, representation, and interpretation skills. We hope that by incorporating our graphing materials into this type of laboratory setting, along with the CAM instructional approach, we will engage and educate students about data and graphing.

5.6.3 Role of Instructors in the Course

The main instructor of the physiology lecture and lab course, Stephanie M. Gardner (SMG), began each lab with a short 10 minute presentation to outline the objectives of the lab for the day, to connect physiology concepts learned in lecture into practice in the lab, and to educate students about the equipment and techniques for lab that week. The graduate student instructor, Aakanksha Angra (AA) presented a formal mini lecture on graphing during week 2 (see Appendix 5A). This mini lecture presented the importance of creating good graphs, introduced students to the three graphing materials, and trained students on using the graph rubric to critique graphs. General feedback was given to students on their graph critiques, advantages, and disadvantages during weeks 8, 11, and 13 (See Figure 5.1). AA also helped SMG by demonstrating equipment use, data acquisition, and analysis using software designed specifically for physiology data acquisition. Besides the short lectures and demonstrations, the instructors functioned as facilitators and advisors to the student groups as they engaged in open-inquiry labs. If students posed questions relating to data transformation and visualization, AA and SMG began each interaction by guiding students through the step-by-step guide (See Table 4.3), i.e. prompting students to state their research question, hypotheses, and variables measured, before deciding on a graph type. If students struggled with visualizing their graph in Microsoft Excel or Origin Pro, the instructors asked students to sketch a graph with pen and paper and verbally articulate their reasoning with their graph choice and data displayed. The instructor then taught students how to organize data in the graphing software in order to create their desired graph. Field notes on the type of feedback instructors provided to the students were kept and discussed at weekly meetings, to ensure uniformity in providing feedback.

5.6.4 Tracking Usage of Graphing Materials

All three graphing materials were available to students throughout each of the intervention semesters via Blackboard Learn. The statistics tracking feature was enabled in Blackboard Learn, with allowed us to document when and how often each student downloaded the graphing material from Blackboard Learn.

5.6.5 Application of the Graph Rubric

The graph rubric (see Figure 4.4) was the most commonly used material in the physiology course during the intervention semesters. It served three purposes. First, the graph rubric was used by the instructors to provide students graded feedback on graphs they constructed and was part of the graph presentation (See Appendix 5C); second, the rubric was used by the students in the physiology lab to critique each other's graphs (See Appendix 4B for an example). After each presentation in the semester, AA gathered the PowerPoints from each group, if multiple graphs were constructed by the group, AA randomly chose one graph from each group, and pasted the selected graph, along with the research question and hypothesis, and a cryptic but informative group number for organizational purposes and to maintain anonymity of the groups into PowerPoint. The graphs were not adjusted to size or modified in any way in order to preserve the original graph. Each graph with its corresponding information was printed on half a sheet of paper, using a colored printer, again, to preserve the original intended purpose by the group. The half sheet of graphs were organized so that no one in the same group received their own graph to critique, and each group member received a different graph, to encourage independent thinking and critique of the graph. Students filled out the graph rubric digitally and uploaded it to Blackboard Learn, with the unique graph identifier for organizational purposes. Unlike the graph rubric from which students received feedback from the instructors, the graph rubric that students used to critique graphs did not have a grading scale or points associated with the graph categories. Students were asked to critique the graph and to provide detailed and constructive comments justifying their score. Students did not receive formal feedback on their critiques but were monitored for efforts. These graph critiques were worth 5 points. The third reason for using the graph rubric was because it aligned the research objective with our intervention and past research confirms that this technique is useful when conducting interventions (Brown, 1992; Anderson & Shattuck, 2012).

Students critiquing peers' graphs instead of their own was a modification from the non-intervention semesters because we wanted to expose students to other types of graphs, and give them practice critiquing graphs, which is part of MRC.

5.6.6 Pre and Post Surveys

Students enrolled in the physiology laboratory during the intervention semesters were given a pre and post survey to evaluation the impact of the graphing materials. The pre-survey was given to the students at the end of the first lab in Week 1 and the post-survey was given to the students at the end of the last lab in Week 16. A cross-over design was used where half of the students in the lab section were given the bacteria scenario and the other half was given the plant scenario for the pre-survey (Piantadosi, 2013). For the post-survey, the scenarios were reversed.

The scenario in the pre and post graphing surveys and prompts were used from our previous study with experts and novices in Chapter 2. Please consult section 2.5.1 Development of the Graphing Scenario in Chapter 2 for details. The only modification we made was that in the pre and post surveys, we did not ask students to construct a graph, but rather indicate how they would organize and label the axes and the type of data they could plot in their graph.

Although we wanted to initially include a graph construction portion in the survey, we decided to omit this due to time constraints and potential frustrations exhibited by students using Microsoft Excel. There were two reasons for presenting students with the bacteria and plant scenarios. The first reason was that these scenarios are very general and would give all students the same advantages. If we had used a physiology scenario, we could have unknowingly put students who had no prior physiology knowledge at an unfair disadvantage. For the purpose of this study, we were interested in students' reasoning with graph choice and construction and not their real or perceived physiology knowledge needed for the task. The second reason was that these scenarios were previously validated in our expert-novice interviews, so we knew that these scenarios and prompts were appropriate when trying to understand student reasoning with graphs.

Also in the post survey, we asked students to indicate their confidence with graph construction and evaluation on a three point Likert-scale and to rate the usefulness of the graphing materials. We also asked students to share any comments or suggestions they had regarding the graphing materials. Questions asked in the pre survey can be found in Appendix 5D and questions from the post survey can be found in Appendix 5E.

5.7 Data Analysis and Coding

5.7.1 Pre and Post Graphing Surveys

In order to analyze the pre and post survey data, we utilized the same coding scheme as previously reported in Chapter 3. Please consult section 3.7.1 Student Reasoning with Graph Choice in Chapter 3 for details. The reason that we chose to use the coding scheme developed in Chapter 3 to analyze student pre and post reflections instead of using the coding scheme developed in Chapter 2 was because the method of data collection was through written responses and we wanted to ultimately compare student reasoning with their graph choice between the non-intervention and intervention semesters. In Chapter 2, the method for collecting data was through think-aloud interviews, with an interviewer present to gently probe the participant to articulate their reasoning where necessary, and so different themes may have emerged. Future research can re-code written student reflections from the pre and post survey with the coding scheme derived from the expert-novice think-aloud interviews.

5.7.2 Evaluation of Student Graphs using the Graph Rubric

To analyze the quality of the graphs constructed by students over the non-intervention and intervention semesters, graphs were removed from their PowerPoint presentations, along with the research question and hypothesis and pasted into a Microsoft Excel file, where they were individually scored using the graph rubric. The points assigned to each graph rubric category were weighted depending on the level of difficulty decided by SMG, AA, and discipline-based education scientists part of the Purdue International Biology Education Research Group (Appendix 5B). For example the point values within the graph mechanics category were lower than those in the graph choice category because labeling of axes is not of the same cognitive engagement level as evaluating graph types for the data. Please refer to the graph rubric in Appendix 5B. We report these data in Figures 6 and 7, when showing the range of the graph rubric scores. However, since we also wanted to look at the graph rubric category, alignment separately, we separated this out from the rubric. Data that show graph alignment are displayed in Figures 13 and 14. In Figures 8,9,10, and 11, we report the average of the graph scores for each of the graph categories, the mechanics, communication, and graph choice, which excludes the subcategory, alignment. A Kruskal-Wallis ANOVA was performed for data from each of the graph categories using SPSS (SPSS, V.22). Inter-rater reliability was performed between AA and SMG on 20% of the graphs.

5.7.3 Measuring Alignment

The alignment category is the last subcategory featured in the graph rubric. This refers to choosing and constructing a graph that displays the data in a manner that facilitates the answering of the research question or evaluation of a hypothesis. Student groups in both the non-intervention and intervention semesters were encouraged to make exploratory graphs of their data because it helps to visualize the patterns and trends showcased by the data. However, students were informed that at least one graph that they used in their PowerPoint presentation should be aligned with their research question and hypothesis. When we modelled the graphing materials to the students, we discussed the importance of alignment. We also implemented this in our instruction. Whenever we walked around the classroom offering help to the students, we asked them to reiterate their research question, hypothesis, and the variables they were manipulating and those that they were measuring. Because some graphs created by students might be used for data exploration or designed to highlight additional data trends, the evaluation of alignment was not conducted on all graphs. We did not want to penalize the student groups who made exploratory graphs to understand their data. Therefore, we report the results of the graph alignment separately in our results and discussion.

5.8 Results

The overarching objective of this study was to implement the step-by-step guide, guide to data displays, and graph rubric in a classroom setting to see if they have a positive impact on students' graph reasoning and construction. To accomplish this objective, we sought to answer the following questions:

Does the application of the Cognitive-Apprenticeship Model, (CAM) in an upper-level undergraduate physiology laboratory with associated instructional materials and approaches, improve students' graph choice and construction over the intervention semesters as compared to non-intervention semesters?

1a) Does consistent use and targeted feedback with the graph rubric improve graph mechanics, communication, choice, and alignment with the research question and hypothesis during the non-intervention and intervention semesters?

1b) Does usage of the guide to data displays improve students' awareness of various graph types over the course of the semester?

5.8.1 Graph Choice in the Pre and Post Surveys

In order to understand the types of graphs suggested by students for the pre and post graphing surveys for both intervention semesters, we report findings in a series of pie charts in Figure 5.2. We report the percentage of graphs constructed in the semesters to show that regardless of the semesters, the same trends appeared across graph choice. In the pre-survey, line graphs were the most common types of graphs created in both semesters. In Spring 2015, bar graphs were the second common graphs, followed by scatter and then box and whisker. In the Spring 2016, scatter plots were the second common, followed by bar graphs, and then histograms. In the post-surveys, line graphs remained a popular choice during both Spring 2015 and 2016 semesters, followed by bar graphs as the second choice. In the post survey we discovered that students suggested a larger diversity of graph types compared to the pre survey. For instance, in the post survey for Spring 2015, we see 12% of the students suggest box and whisker plots, while 3% suggest histograms- a graph that was not present in the pre survey. In the post survey results for Spring 2016, 9% of the students suggested box and whisker plots, 4% suggested dot plots, and 2% suggested stem and leaf plots.

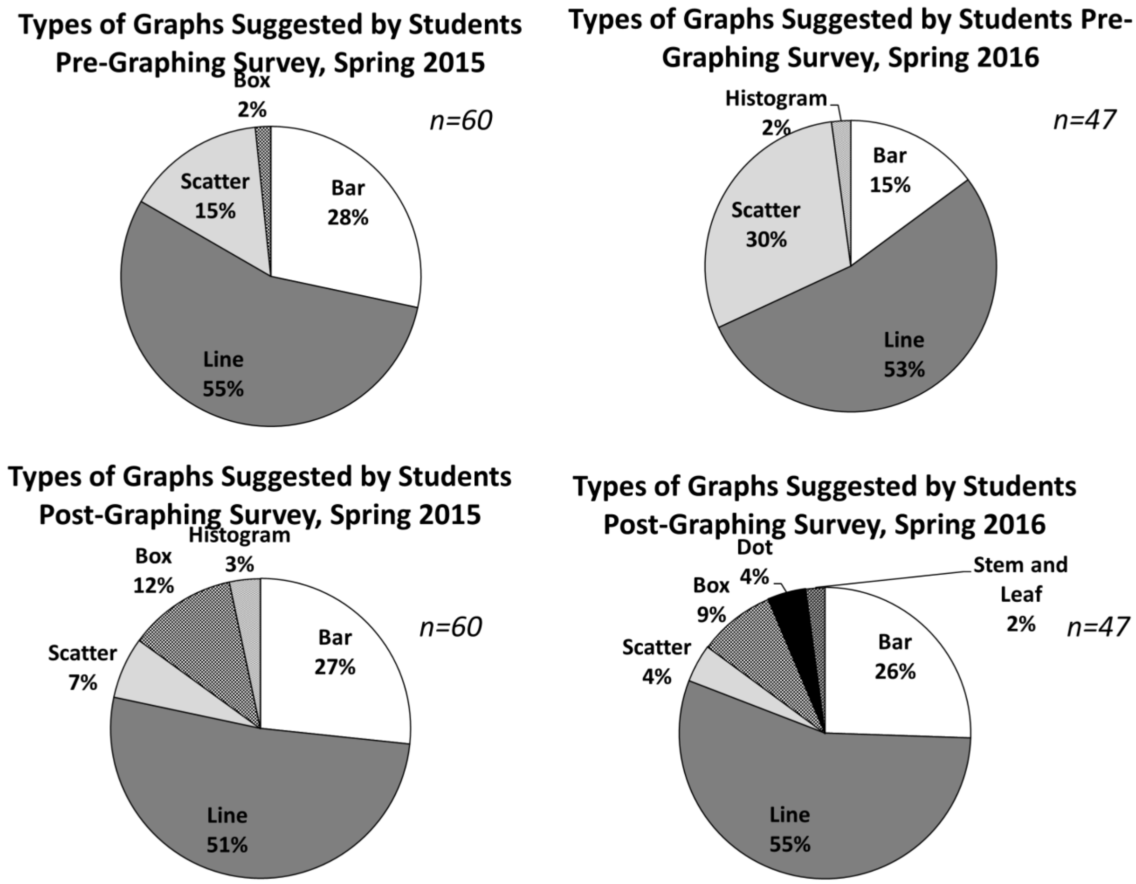


Figure 5.2: Types of graphs suggested in the pre and post survey: The series of pie charts displays the graph choice for the pre and post surveys for each intervention semester. The most common graph choice was line graphs in both pre and post surveys. In the figure above, *n* indicates the number of students.

5.8.2 Reasoning with Graph Choice

To assess a change in students' reasoning with graph choice, we report pre and post survey data for the prompt, "Why did you choose this graph type over others?" from 60 students in Spring 2015 (See Figure 5.3) and 49 students in Spring 2016 (see Figure 5.4). Students' pre and post survey responses were mapped onto the coding scheme reported in Chapter 3 (Appendix 3D). Figure 5.3 displays the themes that emerged for graph choice in the form of stacked bar graph with graph type and sample size on the x-axis and percentage of themes on the y-axis. The percentage of themes were calculated by first counting the number of instances of a code that appeared for each theme, within the year, dividing it by the total occurances of themes for that year, and multipying it by 100.

Student responses fell mainly within the themes of *graph interpretation* and *communication* for both semesters. Although we expected students to be more reflective and mention *experimental concepts*, in the post surveys, we did not see see this as a common theme.

Since the themes *graph interpretation* and *communication* were the most common themes, we looked into each one at a finer scale. Appendix 5F shows the primary codes for the themes *graph interpretation* and *communication.*

Looking broadly at the primary codes for *graph interpretation,* we see a greater presence of the primary code comparison, followed by differences and change in both the pre and post surveys in both intervention semesters. The primary code, *relationship* was the least common across all different types of graphs. The primary code, *take-home message* was common in the post-survey for spring 2016 for line graphs and post survey for scatter plots in Spring 2015. However, there was a small sample size of students who suggested scatter plots.

There were three primary codes that were prevalent within the theme of *communication* (Appendix D). The most common primary code was *function*, followed by *figural effects*, and then *graph components.* The primary code, *figural effects* (diSessa, 2004), states that students pay attention to aesthetics and Gestalt rules such as proximity of lines and offer their personal opinion on the quality of the graph, instead of focusing on the function of the representation. The primary code *graph components* (Åberg-Bengtsson, L., & Ottosson, 2006;Schriger & Cooper, 2001; Cooper et al 2002; Kosslyn, 1994; DiFazio, 1990; Elliot et al., 2006; Leonard & Patterson, 2004; Patterson and Leonard 2007; Ainley 2000; Brasell & Rowe 1993) consists of three secondary codes, *variables, describing the layout of the graph,* and *linking the data plotted to the graph.* The primary code graph components was a common theme for the pre survey in spring 2015 for bar and line graphs. Graph components was also common in pre and post bar graphs for spring 2016 and in the pre survey for scatter plots in Sprong 2016.

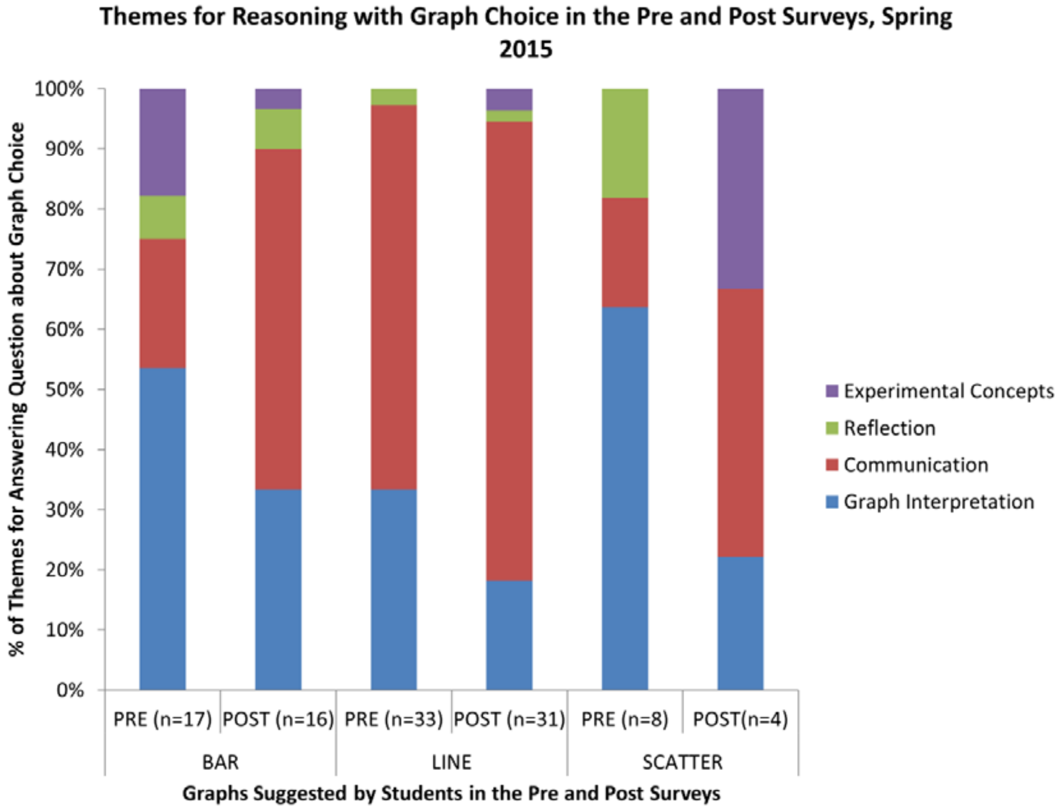


Figure 5.3: Themes for reasoning with graph choice for Spring 2015: Types of reasoning associated with graph choice for pre and post surveys in Spring 2015 and the percentage of themes associated with each graph choice.

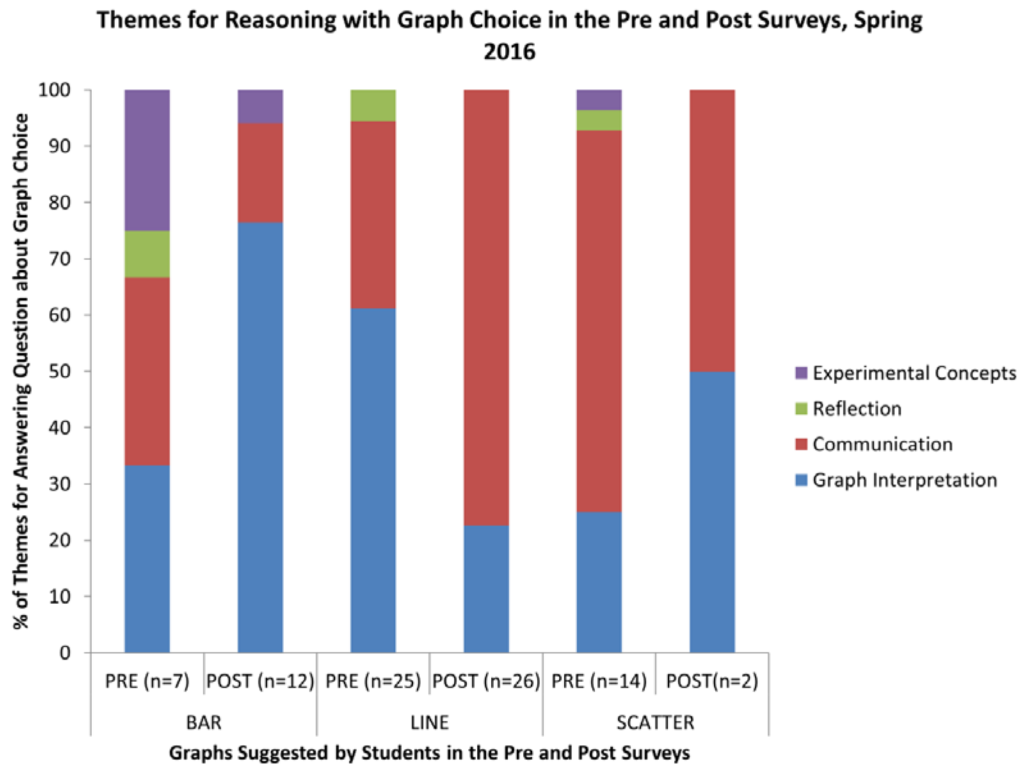


Figure 5.4: Themes for reasoning with graph choice for Spring 2016: Types of reasoning associated with graph choice for pre and post surveys in Spring 2016 and the percentage of themes associated with each graph choice.

5.8.3. Attributes of Graphs Constructed over the Non-Intervention and Intervention Semesters

Students had 5 opportunities over the course of the semester to gather and analyze data, summarizing them in graphs for oral presentation. The types of graphs created by student groups from both the non-intervention and intervention semesters are illustrated in a series of pie charts in Figure 5. In both the non-intervention and intervention semesters, bar graphs were the most common type of graph constructed. It is also interesting to note that dot plots and box and whisker plots appeared in the first lab, in the intervention semesters, but appeared later in the non-intervention semesters. There was a higher percentage of bar graphs in the beginning of the non-intervention semesters and it gradually decreased (46% to 27%), whereas in the intervention semesters, the percentage of bar graphs increased from 45-52%. The percentage of scatter plots decreased from 46 to 31% in the non-intervention semesters and went from 30% to 0% in the intervention semesters. The patterns for line graphs constructed were very similar across semesters and labs, with a gradual increase observed in each progressing lab.

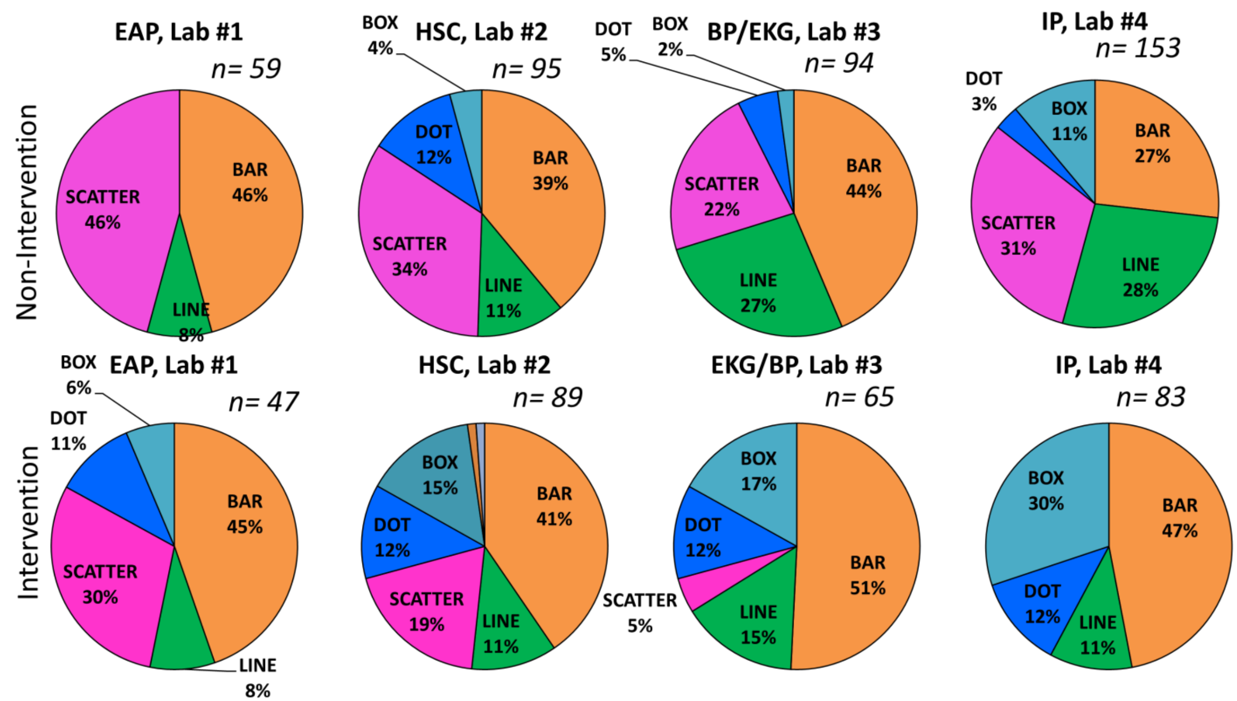


Figure 5.5: Types of graphs constructed in the non-intervention and intervention semesters: The series of pie charts displays the types of graphs constructed for each lab and for each semester. The most common graphs constructed across all years were bar graphs. It is interesting to note the early appearance of dot plots and box and whisker plots in the intervention semesters. In the figure above, *n* indicates the number of total graphs created for that lab.

5.8.4 Graph Quality over the Non-Intervention and Intervention Semesters

In order to see if the graphs constructed by student groups improved over the semester, the graph rubric was used to assess the student graphs (Appendix 5B). In Figures 5.6 and 5.7, we report the range of graph rubric scores over the non-intervention and intervention semesters for bar and line graphs only because these were the only graph types that were present in both semesters and across all four labs, which allows us to adequately compare data between the two years. For the bar graphs, the average of the graph scores did not change very much over the course of the semester in the non-intervention semesters, but increased in the intervention semesters, with majority of bar graphs scoring close to maximum points possible. For the line graphs, the average graph scores for the intervention semester for the first lab were below that of the score in the non-intervention semester. By the end of the semester, the average score from 11 line graphs in the intervention semester was higher than the average range of graph scores from the 34 graphs from the non-intervention semester.

In order to examine if there were differences in particular graph attributes, we examined the scores within each of the large graph rubric categories, graph mechanics, graph communication and graph choice. In Figures 5.8,5.9,5.10, and 5.11 we show the graph scores broken down by the large graph rubric categories. For the first lab, A Kruskal-Wallis ANOVA showed a significant difference ($p<.05$) in graph mechanics and graph choice between the non-intervention and intervention semesters. For the second lab, A Kruskal-Wallis ANOVA shows a highly significant difference in graph mechanics and communication ($p<..005$ and $p<.0005$, respectively) between the non-intervention and intervention semesters.

For the third lab, A Kruskal-Wallis ANOVA shows a significant difference ($p<.05$) in graph mechanics and highly significant difference ($p<.005$) in graph choice between the non-intervention and intervention semesters. For the last lab, A Kruskal-Wallis ANOVA shows a highly significant difference in graph mechanics, communication, and graph choice between the non-intervention and intervention semesters.

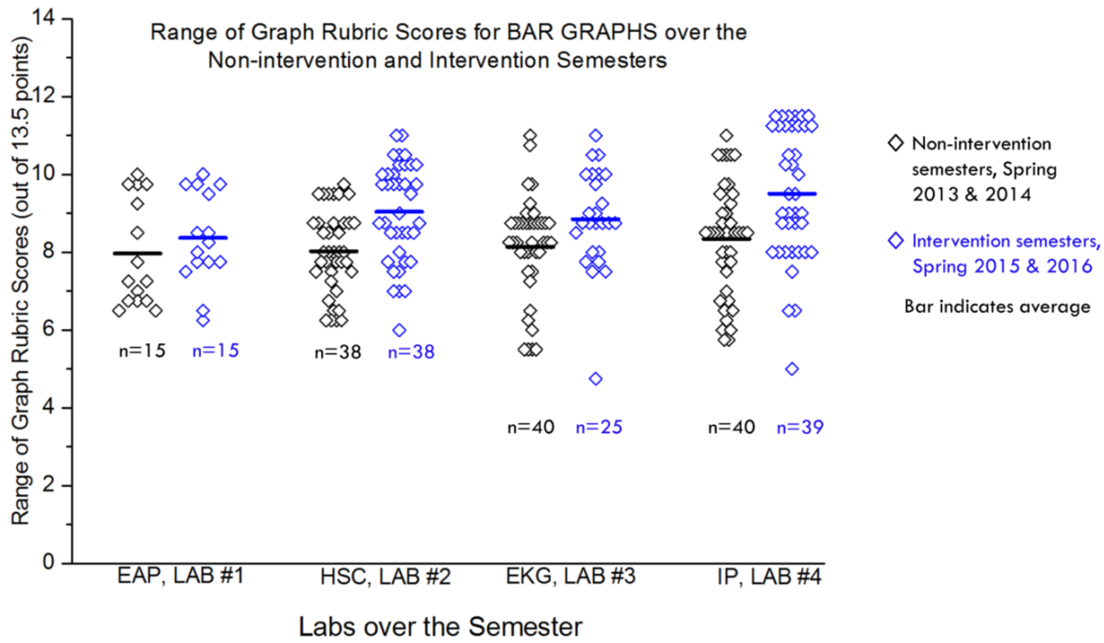


Figure 5.6: Range of Graph Rubric Scores for Bar Graphs: The dot plot shows the range of graph rubric scores over the course of the non-intervention and intervention semesters. The average of the graph scores did not change very much in the non-intervention semesters, but increased in the intervention semesters, with majority of bar graphs scoring close to maximum points possible. The number of graphs created for each lab for the semester of interst are indicated by *n*.

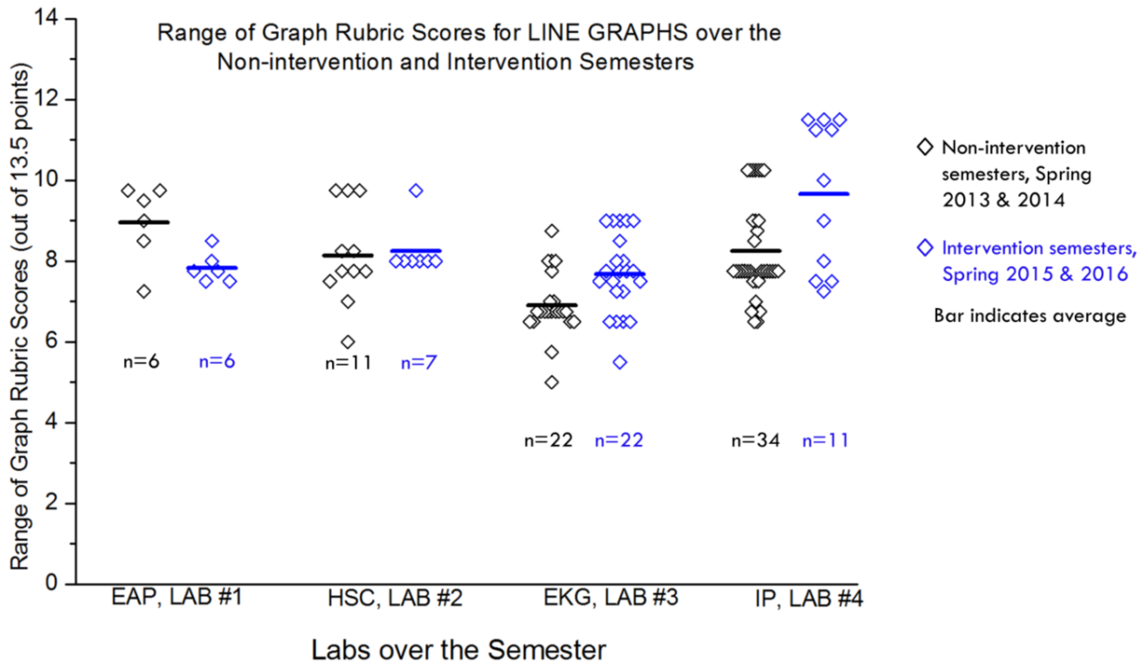


Figure 5.7: Range of Graph Rubric Scores for Line Graphs: The dot plot shows the range of graph rubric scores over the course of the non-intervention and intervention semesters. The average graph score for the intervention semester for the first lab were below that of the score in the non-intervention semester. By the end of the semester, the average from 11 line graphs in the intervention semester were higher than the average range of graph scores from the 34 graphs from the non-intervention semester.

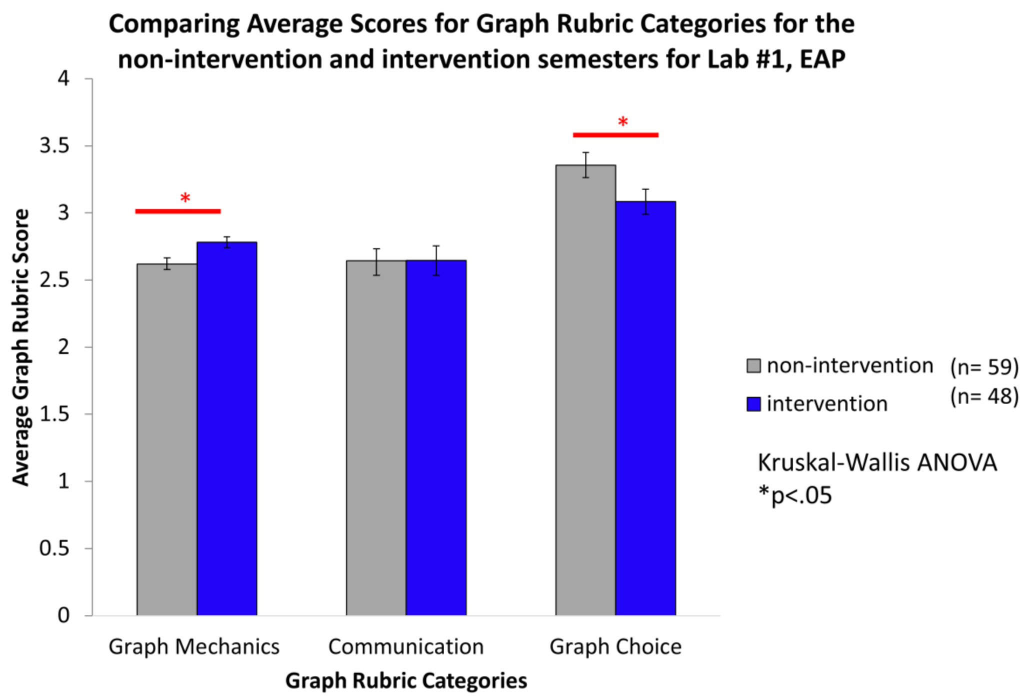


Figure 5.8: Average Scores for Graph Rubric Categories for Lab #1: This graph compares all graphs constructed in the non-intervention and intervention semesters, broken down by their mechanics, communication, and choice. A Kruskal-Wallis ANOVA shows a significant difference ($p<.05$) in graph mechanics and graph choice between the non-intervention and intervention semesters.

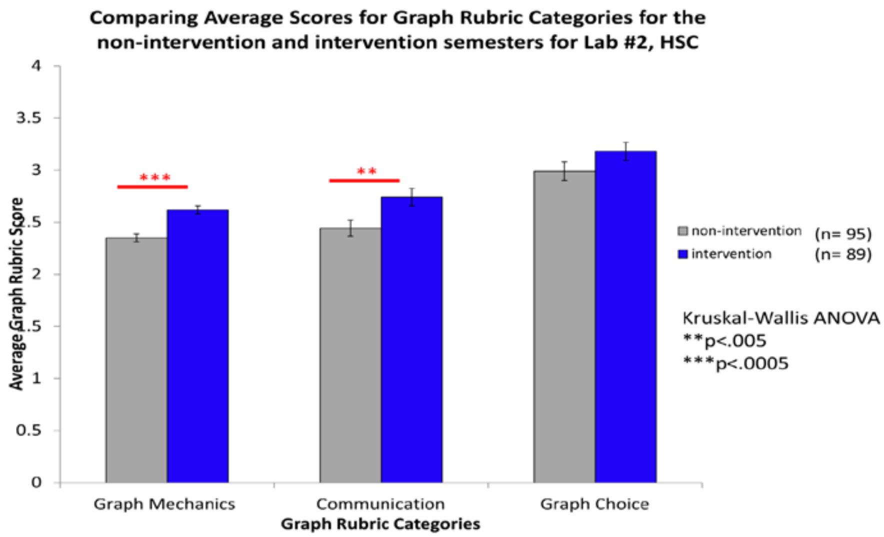


Figure 5.9: Average Scores for Graph Rubric Categories for Lab #2: This graph compares all graphs constructed in the non-intervention and intervention semesters, broken down by their mechanics, communication, and choice. A Kruskal-Wallis ANOVA shows a highly significant difference in graph mechanics and communication between the non-intervention and intervention semesters.

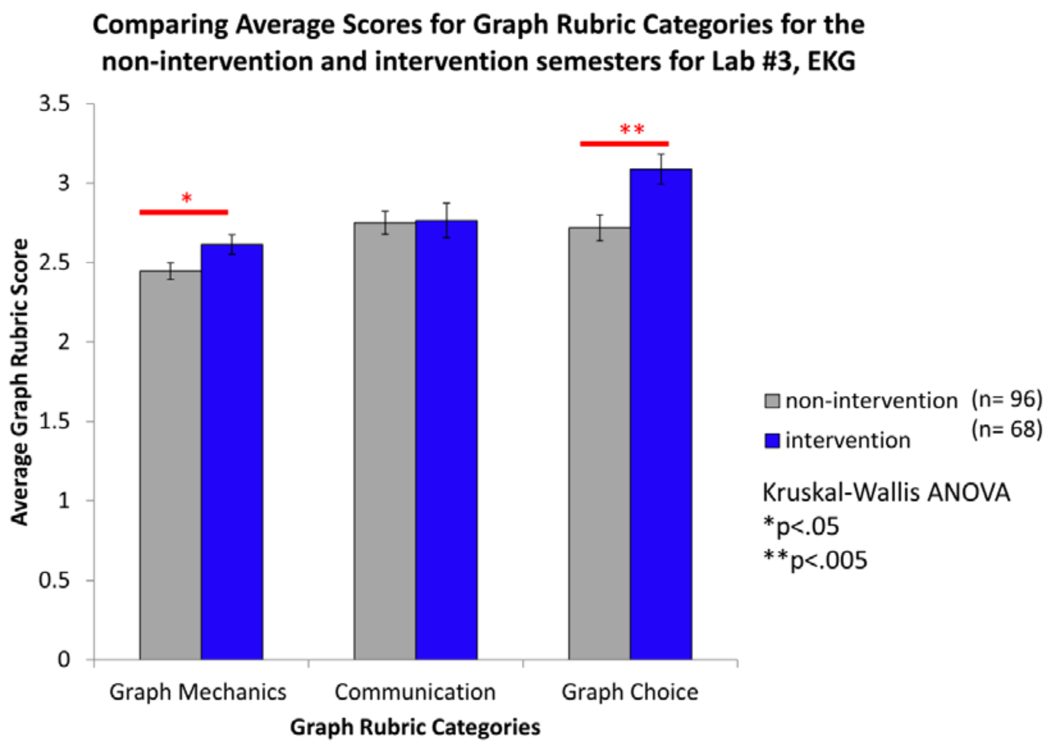


Figure 5.10: Average Scores for Graph Rubric Categories for Lab #3: This graph compares all graphs constructed in the non-intervention and intervention semesters, broken down by their mechanics, communication, and choice. A Kruskal-Wallis ANOVA shows a significant difference ($p<.05$) in graph mechanics and highly significant difference ($p<.005$) in graph choice between the non-intervention and intervention semesters.

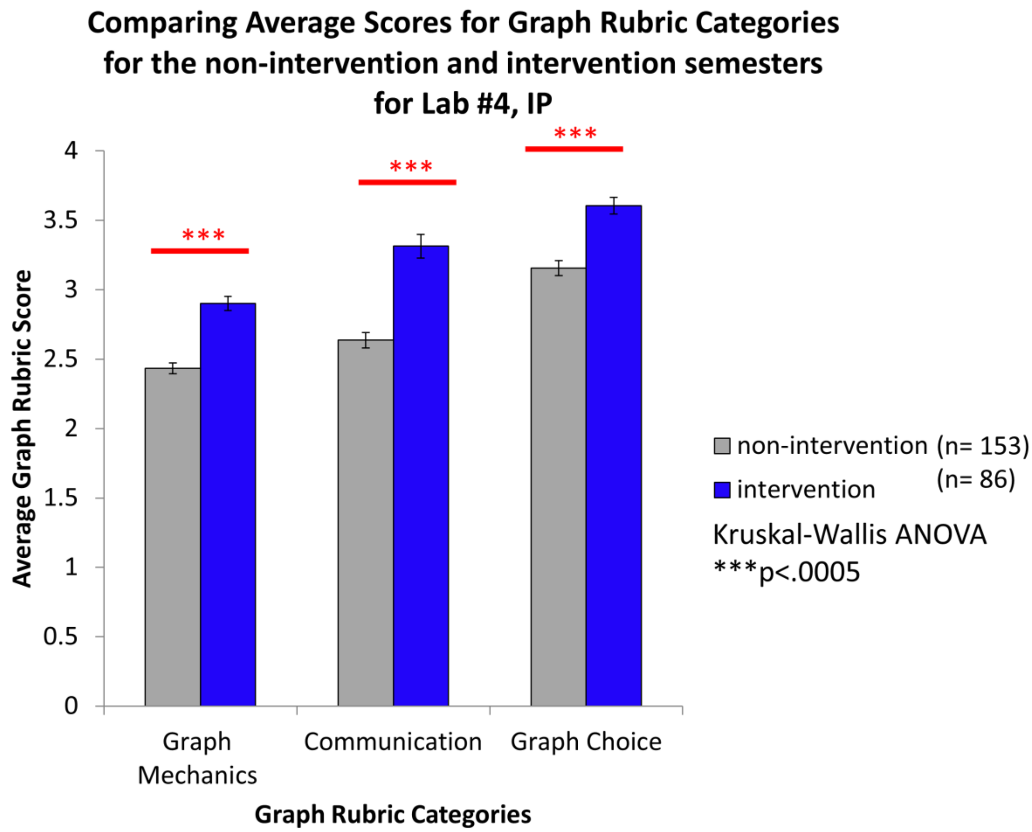


Figure 5.11: Average Scores for Graph Rubric Categories for Lab #4: This graph compares all graphs constructed in the non-intervention and intervention semesters, broken down by their mechanics, communication, and choice. A Kruskal-Wallis ANOVA shows a highly significant difference in graph mechanics, communication, and graph choice between the non-intervention and intervention semesters.

5.8.5 Alignment of Graphs to Research Question and Hypothesis

One important function of graphs is the display the data in a way that allows one to answer their research question and/or test their hypothesis. This attribute is evaluated within the graph choice category of the graph rubric and over all of the non-intervention and intervention semesters, student groups were asked to align their graphs to their research question and hypothesis of interest in their study. Figure 5.12 shows the findings from the non-intervention semester and Figure 5.13 shows the findings from the intervention semesters. We see that the non-intervention semesters, starting with the first lab, the majority of the student groups' graphs (22 out of 35 groups showed partial alignment). After giving students feedback on the graphs constructed for the first lab, we see an increase improvement in alignment with students' graphs in the second lab, where 21 out of the 35 groups showed complete alignment. For the third lab in the non-intervention semester looking at the EKG and blood pressure and the last lab in the non-intervention semester, while we discovered more groups completely aligned their graph to the research question and hypothesis, equally as many student groups had graphs that were partially aligned. Examples of graphs showing complete, partial, and no alignment can be accessed in Appendix 5G.These data are displayed in Figure 5.12 below.

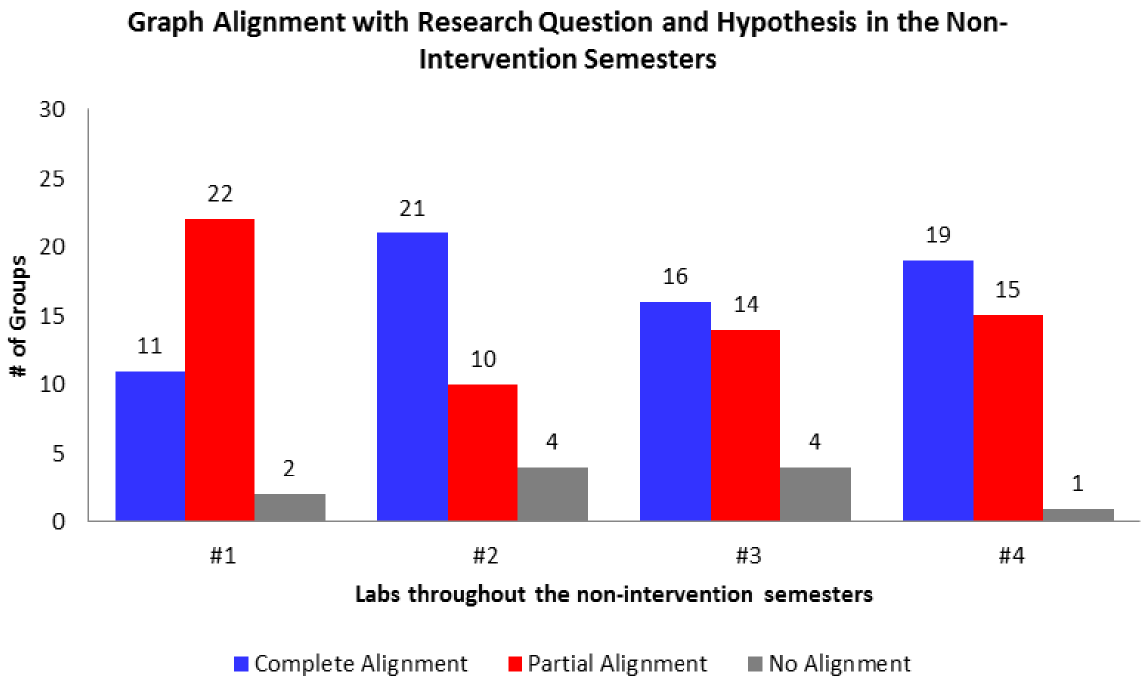


Figure 5.12: Graph Alignment with Research Question and Hypothesis for the Non-Intervention Semesters. This graph shows that by the end of the semester, 19 groups of students had at least one graph that aligned with the research question and hypothesis and 15 groups of students had at least one graph that had partial alignment with the research question and hypothesis.

In the intervention semester, student groups started out with the same general trend in the first lab, with 21 groups of the 34 total had at least one graph that was partially aligned and 12 of the 34 groups had at least one graph that was completely aligned (Figure 5.14). After feedback on the first lab, there was a slight increase in graph alignment for the second lab, with 6 more student groups falling into this category. We see a slight decrease in alignment for the third lab, with only 14 groups falling into the perfect alignment category and 18 groups falling into the partial alignment category. In the last lab however, we discovered that 28 of the 34 groups completely aligned their graph to their research question and hypothesis and only 5 groups showed partial alignment and 1 group showed no alignment. . Examples of graphs showing complete, partial, and no alignment can be accessed in Appendix 5G.

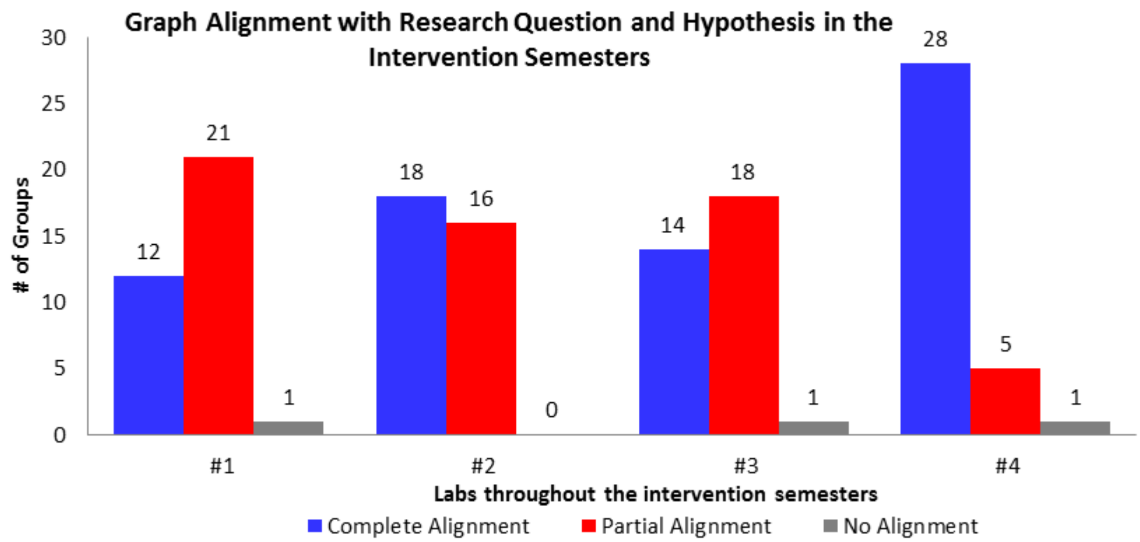


Figure 5.13: Graph Alignment with Research Question and Hypothesis for the Intervention Semesters: This graph shows that by the end of the semester, 28 groups of students had at least one graph that aligned with the research question and hypothesis.

5.8.6 Student Rating of Graphing Materials and Percentage of Downloads

In addition to examining the impact of the instructional materials used within the CAM pedagogy on student graph choice and construction, we wanted to understand how students perceived the graphing materials and activities and whether or not they found these materials useful. Students were asked to respond to a series of questions on the post-survey on Qualtrics. On a Likert scale, students were asked to rate the usefulness of the graph rubric, the step-by-step guide, the guide to data displays, and peer evaluations. The majority, approximately 80% of the students found the graphing materials and the peer evaluations to be useful and conducive to their learning (Figure 5.14). When students were asked to evaluate their confidence with graph construction and evaluation, between 78 -90 percent of the students indicated that they felt confident with graph construction and evaluation (Figure 5.15). Around 20 percent of the students indicated that they felt somewhat confident with these skills and about 5 percent of the student in Spring 2016 indicated that they did not feel confident with graph evaluation (Figure 5.15).

Data from Blackboard's Statistics Tracking Function reveals that students downloaded the step-by-step guide, guide to data displays, and the graph rubric over the semester, with the most frequent downloads occurring on Tuesdays and Thursdays, when students were in the physiology laboratory. In both intervention semesters, the step-by-step guide and guide to data displays were equally the most downloaded resources. When students were asked to comment or provide suggestions regarding the usage of these materials, we saw some students state their mild frustrations with computer graphing software. For example one student stated that "*Origin is very helpful once you understand how to use it! But the current system is very effective and I would not change much if anything.*" Another student suggested: "*Maybe some video clip links for better understanding of the materials would be useful.*" and another student suggested, **"***I wish we would have had a lab session where we went into more detail about making certain graphs and making statistical analysis like one-tailed and two-tailed t-test, and ANOVA.*" Students also commented on the graphing materials and their confidence with graphing in general:

- *"I feel like I learned a lot this semester about graphs. The rubrics helped out a lot."*
- *"I definitely think I am more confident in my graphing abilities because of this semester and am more detailed when making them."*
- *"I certainly learned how important it is to incorporate certain elements into a graph! I cannot believe how much I was missing before."*
- *"I believe I have improved with graph formation, but still am not completely comfortable in knowing which to use right off the bat."*
- *I'm confused when to used average data vs. raw data.*
- *I learned a lot, but it still seemed like all we ever did were line or bar graphs. I had hoped to get more experience with more "grown up" ways of presenting data.*

In Appendix 5H, we share a complete list of students' thoughts on how these materials helped or did not help them learn.

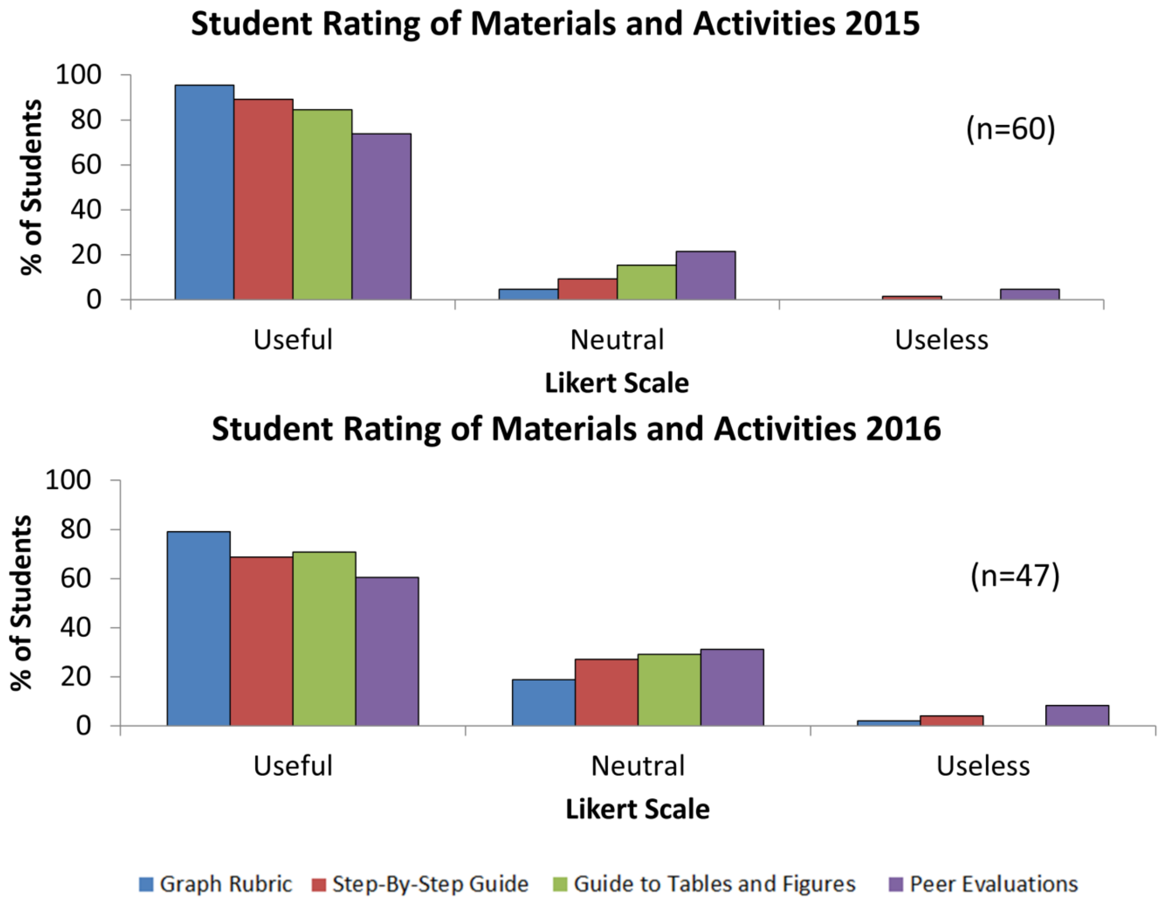


Figure 5.14: Student Rating of Materials: These bar graphs show how students in Spring 2015 and 2016 rated the three graphing materials and how they perceived peer evaluations of graphs. Most students found the materials and peer evaluation of graphs to be a useful activity.

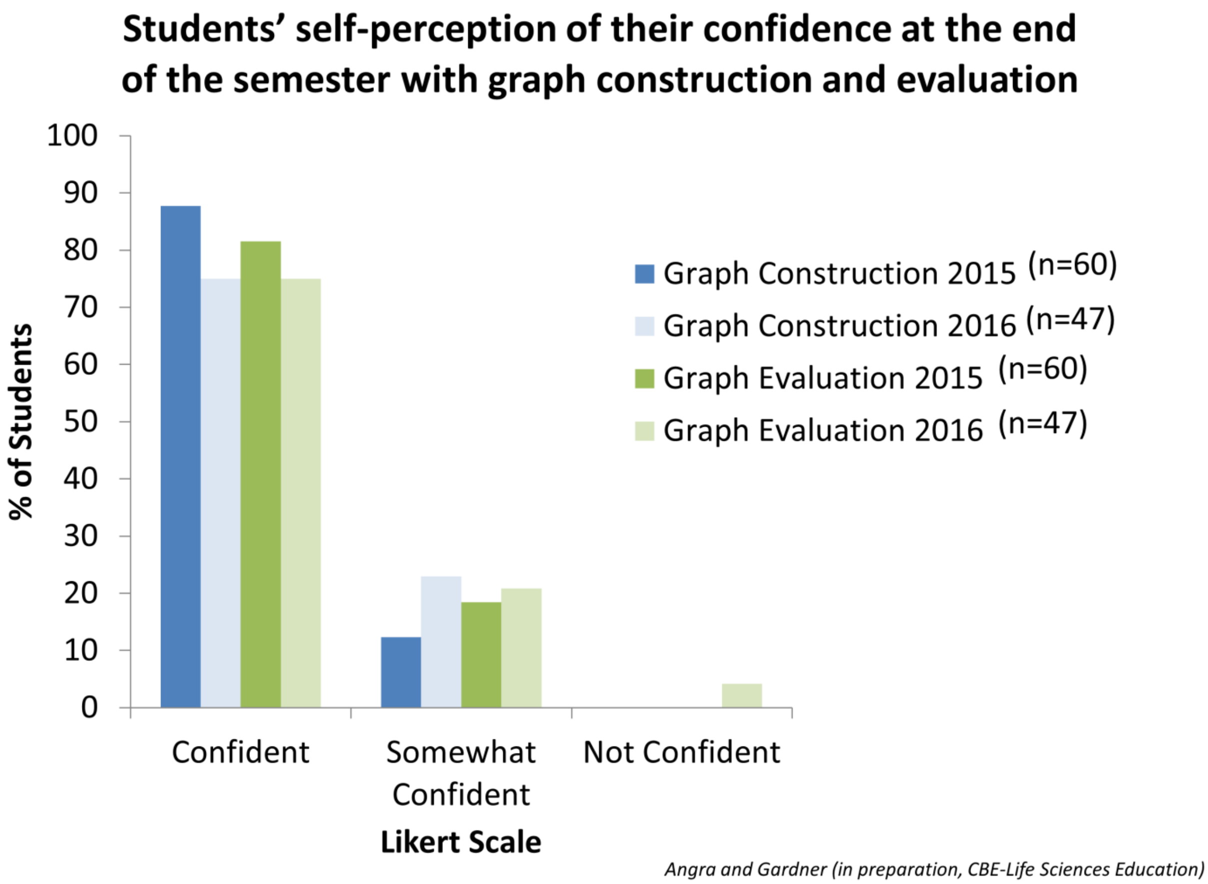


Figure 5.15: Students' self-perception of their confidence at the end of the semester with graph construction and evaluation: In both the intervention semesters, between 78- 90% of the students felt condident with graph construction and evaluation. Approximately 20% of the students felt somewhat confident in their skills and about 5 % of the students in the Spring 2016 semester did not feel confident with evaluating graphs.

5.9 Discussion

Recommendations for undergraduate biology education include engaging students in the practices of science as a critical feature of the learning of their discipline (Brewer and Smith 2009). One of the fundamental practices for science and conducting research is working with data, which consists of understanding, consolidating and communicating data in graphic representations. Past studies have shown student difficulties with different aspects of graph choice and construction at the K-12 and undergraduate levels (Leonard and Patterson, 2004; Tairab and Al-Naqbi, 2004; Padilla, McKenzie and Shaw, 1984; Brassell and Rowe, 1993; Ainley, 2000; Bray-Speth et al., 2010; McFarland, 2010; Angra & Gardner, 2016a; Angra & Gardner, 2016b).

In this study we implemented three evidence-based graphing materials in an upper-level undergraduate inquiry-based physiology laboratory classroom. We utilized these materials using the recommendations made by the Cognitive Apprenticeship Model (CAM) pedagogy to see if there was an improvement in students' graph choice and construction over the intervention semesters as compared to non-intervention semesters. Particularly, we wanted to understand if the continuous usage of the graph rubric improved students' graph mechanics, graph communication, graph choice, and the alignment with the research question and hypothesis. We also wanted to observe if the usage of guide to data displays improved students' awareness of various graph types over the course of the intervention semester. Our findings suggest that compared to the non-intervention semester, implementation of the graphing materials under the CAM did improve students' graph choice and construction over the course of the intervention semesters. Figures 5.8-5.11 display scores for the large graphing categories over the four labs in the semester. By the end of semester, graphs constructed by students in the intervention semesters were significantly better (Kruskal Wallis ANOVA, $p<.005$) than the graphs constructed during the non-intervention semesters. The reason for visible improvements in student graphs can be explained by the fact that our graph rubric acted as an assessment and learning tool that aligned with our research objectives, learning objectives for the students, and instructional activities. Past research confirms that this technique is useful when conducting interventions (Brown, 1992; Anderson & Shattuck, 2012). Additionally, rubrics are valuable tools in the classroom because they make learning goals explicit for both the instructors and students (Allen and Tanner, 2006; Wolf & Stevens, 2007).

Although we did not monitor student use of the step-by-step guide or the guide to data displays, we did refer to these scaffolds when coaching students in the laboratory. We noticed immediate improvements in students' graph mechanics, and choice in the first lab of the intervention semester, when compared to the first lab in the non-intervention semester (Figure 5.8). Graph alignment to the research question and hypothesis for the lab did not vary much from the non-intervention semester. This could be because it is a difficult skill to master that even professionals have difficulty with (Rougier et al., 2013). We also noticed that in the first lab in the intervention semester, students chose to display their data in dot and box and whisker plots. These graph choices did not appear until the second lab in the non-intervention semester. Even though this is a favorable outcome from the *invention* point of view (MRC, diSessa & Sherin, 2000), it raises concerning questions about students' awareness of the suitability of graph type to their data. In the first lab of the semester, where students spent more time familiarizing themselves with the equipment used to collect data than actually collecting data for the experiment, all student groups had a very small sample size (n=2) with one trial. For this first lab, the 11% of dot plots created (Figure 5.5) were appropriate graphing representations, especially with small data sets (Drummond & Vowler, 2011; Weissgerber et al., 2015; Angra & Gardner, 2016), however the 6% of box and whisker plots were not appropriate (Schriger & Cooper, 2001; Angra & Gardner, 2016), mainly because box and whisker plots excel at showing and comparing distributions of large datasets and also display the mean and median (Schriger & Cooper, 2001; Duke et al., 2016; Angra & Gardner, 2016). By the fourth lab in the intervention semester, we noted significant improvements with students' graph mechanics, communication, and choice (Kruskal Wallis ANOVA, $p<.0005$; Figure 5.11) and noted an improvement with graph alignment (Figure 5.14). The last lab in all the non-intervention and intervention semesters was worth 50 points, whereas the first three labs were worth 10 points, each. Although at first it may seem that more points could have factored into student motivation to perform well on their graph construction, we did not see much improvement in graph mechanics, communication, and choice for the non-intervention semesters (Figures 5.8-5.11). This was also evident in the lack of improvement in the overall graph score with bar (Figure 5.6) and line graphs (Figure 5.7). One reason that we did not see much change in graph quality for the graphs constructed in the non-intervention semesters could be because students were constructing their graphs using Microsoft Excel and Origin, both of which have default settings that are not sufficient for communicating findings (Patterson & Leonard, 2004). For instance, in order to label the graph in Microsoft Excel, one needs to indicate the addition of axes labels and title. Also, in order to inform the reader about the data displayed, students must manually add the sample size and number of trials to the key. The second reason that we did not see a change in student graphs could be that the feedback given to the students using the old version of the rubric, which was developed using instincts (Appendix 3C) was not comparable to the evidence-based graph rubric (Table 4.4). In order to measure tangible outcomes, students need appropriate guidance, validated scaffolds, and feedback from evidence-based graphing materials.

When students were asked to rate these graphing materials and classroom activities, the peer evaluations were not favored by some students, and appeared more than once in the "useless" category (Figure 5.14). Additionally, when students were asked to provide feedback on the graphing materials on the post-survey, although some students found peer-feedback useful, some students did not think it was a worthwhile activity for them. As instructors, our goal for the peer feedback was to expose students to various types of graphs by having them actively critique each others' graphs. We hoped that this process would allow students to remember the various categories of the graph rubric so that when students are constructing their own graphs, they can apply what they learned from the peer evaluations to improve their graphs. Although this activity has the potential to be very useful for students, one way it could have improved in the future is to provide students feedback on their peer evaluations. Another way is to help students see the value of the peer-review process and help realize that this is an expectation of the science profession.

Even though the graph rubric was the most used in the classroom and in various contexts, statistics tracking revealed that it was the least downloaded material. One reason could be that students frequently consulted the presentation and graph feedback given by the instructors, so they may not have felt the need to repeatedly download the graph rubric. Another reason could be that since students were asked frequently (almost every other week) to use the graph rubric to give peer feedback, they may have familiarized themselves with the contents of the rubric and may not have felt the need to repeatedly download this material. Additionally, students could have downloaded and printed the graphing materials one time and referred to them numerous times throughout the semester. Therefore, future research should seek ways to understand how students use these graphing materials both in and outside of the classroom.

The pre and post survey results revealed that when asked about the type of graph students would construct, majority chose line graph with 55% in Spring 2015 and 53% in Spring 2016 (Figure 5.2). Graph choice in the post survey did not change very much. The line graph was still a popular choice with 51% of the students choosing it in Spring 2015 and 55% choosing it in Spring 2016 (Figure 5.2). An interesting finding from the post surveys in both semesters was that students suggested a greater variety of graphs, as compared to the pre surveys (Figure 5.2). This suggests that frequent exposure to the guide to data displays, and to students' graphs presented in PowerPoint presentations and graph critiques could have influenced their graph choice. However, this does not mean that a box and whisker plot or a histogram are appropriate data displays for the plant or bacteria scenario (Appendix 2A). Line graphs were also a common choice across all participant groups in the think-aloud interviews (Appendix 2D). All of the expert professors (chapter 2) chose to display data from the plant or bacteria scenario in a line graph (Appendix 2D). The expert professors justified their reasoning with graph choice with the question and/or hypothesis, prior experience, and discipline-based knowledge (Figure 2.7), whereas the undergraduate students in the interview did not (Figure 2.7).

Student responses for graph choice in both the pre and post surveys for both intervention semesters reveal that students did not explicitly state that they aligned their graph to their research question or hypothesis, and despite being enrolled in an inquiry-based lab and conducting experiments, *experimental concepts* was not a major theme that appeared in the reflections. Although one could argue that the reason for the lack of appearance of *experimental concepts* could be due to the fact that students did not collect data in the plant or bacteria scenario, this was also an uncommon theme in the non-intervention semesters, when students were reflecting on graphs from their own experiments.

Students' written reflections on experimental concepts, research question, and hypothesis when reflecting on graph choice, needs to be improved in both intervention and non-intervention semesters. A future study can be conducted to understand the benefits of reflection and how it effects students' confidence with graph construction and evaluation (Figure 5.15), and how it influences the types of concepts students think about when reflecting and critiquing graphs in the classroom and in the media. Although we provided students with many opportunities in the non-intervention semester with self-reflections on students' own graphs (chapter 3) and with having students give peer feedback in the intervention semester, we did not give any feedback in the non-intervention semester and gave general feedback in the intervention semester, because we were limited with time in the classroom. Since graphing was a secondary component in the physiology laboratory, we had to prioritize what we explicitly gave feedback on, although we were always available in the lab and provided students feedback as needed.

5.10 Project Scope and Future Studies

Six main study design features affect the scope of our conclusions. First, data were collected from 123 upper-level students at a single Midwestern United States research university, enrolled in an upper-division physiology course, with its unique curriculum and learning objectives for students. Since we did not test the utility of the graphing materials in other laboratory classrooms with different student demographics and different instructors, the claims presented here are apply only to the 123 undergraduate students enrolled in this course. In order to fully understand the impact of these evidence-based graphing materials, future work should check the utility of these evidence-based graphing materials at various K-12 and undergraduate institutions in various types of classrooms, and with their own diverse student population.

Second, the extent and depth of detail in which these materials were presented were constrained by the previously established lab syllabus. Since these graphing materials were implemented in a physiology laboratory, the main focus for students was to practice and apply the physiology knowledge acquired during lecture over to the laboratory setting. Graphing was a secondary focus. During each lab period, student groups were tasked with numerous learning opportunities (e.g. conducting literature searches, designing experiments, collecting data, physiology problem solving scenarios, etc.) which had to be completed during the 3-hour span of the lab period. Because of these weekly activities, only one short formal lecture was given to the students at the beginning of the semester on graphing. However, once students began conducting experiments, collecting data, and graphing their findings, the instructors circulated around the classroom and reminded students to consult the graphing materials.

Since we wanted to compare the graphs created in the non-intervention semesters with those created in the intervention semesters, we did not make drastic changes to the lab curriculum during the intervention semesters; we wanted to test just how well new instructional materials emphasizing an important skill such as graphing could fit into an already existing curriculum. Our results do show that even with the given time constraints on lecturing about graphing, students improved their graph choice and construction.

Therefore, we strongly recommend instructors to utilize and incorporate these materials into their classroom, as it will not take much time away from the main learning objectives of the course. One of the most important things about implementing these materials is to present them to students in an engaging manner, where students have frequent exposure and practice using these materials. Since most students in the United States learn about bar graphs and pictographs by the second grade in the context of mathematics (NGSS, 2012), they may have a false sense of confidence with graphing, if these graphing materials are presented in a passive manner (McFarland, 2010). One strategy that we and McFarland (2010) used is to have students critique graphs from science journals and textbooks. Not only does this engage students in the activity, but it encourages critical thinking, and plays a role in increasing students' MRC and helping them transition towards expertise (diSessa, 2004).

Third, as with our previous study reported in Chapter 3, we did not focus at the individual or at the group level, nor did we focus on group dynamics or micromanage groups to ensure that every member of the group had the opportunity to construct graphs. In the previous non-intervention semesters, we encouraged students to exchange group roles so that every person in the group got the opportunity to engage in all the different roles (e.g. principle investigator, equipment specialist, data specialist, etc.). Although we utilized the statistics tracking function on Blackboard Learn, to see if students were downloading these lab materials, we do not know if the person responsible for graph construction consulted these evidence-based graphing materials or not, and if so, to what extent. We also did not keep track of how students were using these materials in and outside of the classroom. It is a possibility that some students downloaded the materials once, printed them out for their reference, and then never downloaded the materials again. Since all group members were given access to these materials for the entire semester, we hoped those who downloaded these materials either helped or checked the final graphs created. We suggest future studies can focus at the level of the student group to see how students engage in experimental design, data collection, analysis, and communicating their findings in the form of a graph. We also suggest encouraging each group member to use the step-by-step guide to help them with their graph choice and construction. This will provide all group members with the opportunity to think about the types of graphs available and provide them with opportunities to discuss data and graphical representations.

Fourth, the purpose of the step-by-step guide was to provide a scaffold for students to use to map their existing knowledge of graphing onto a more expert model. Although students were asked to use this guide, due to the preexisting time constraints, we did not require students to fill it out or give them feedback. One limitation of the step-by-step guide is that since it was developed for students at the novice to journeyman level, it may not be as helpful for students who are at the expert graphing level. l. Another limitation is that we do not know if the usage of the step-by-step guide improves student reasoning with graphing. A future follow-up study could conduct end of the semester, think-aloud interviews with students who downloaded and used the step-by-step guide and compare their step-wise graph construction with those of expert professors and graduate students (chapter 2).

Fifth, students were asked to reflect on their peers' graphs four times over the semester. Although we trained students to reflect on graphs using the rubric and discussed how to state the advantages and disadvantages of the graphical representation, we did not give individual feedback to students critiquing their peer's graph. Instead, we coarsely coded student responses for the advantages and disadvantages and provided feedback to the entire lab section during weeks 8 and 11. Considering the purpose of this study, we did not track the quality of reflections for each student over the course of the semester and neither did we associate the accuracy of the graph to the graphing reflection. Future studies can focus on peer critique of the advantages and disadvantages, as well as graphs. One way instructors can help students improve their graph critique is by offering targeted individual feedback. Not only can expert instructor feedback improve students' metacognition, but may also positively impact how students choose and construct their own graphs.

Sixth, we did not account for motivation. Students worked in groups and were expected and encouraged to contribute equally, but field notes suggest that some students were more motivated than others. Other activities could have influenced students' wellness and motivation in the course (e.g. exams and projects in other courses, part-time jobs, extra-curricular activities, personal matters, etc.). Although we did not measure motivation, it was noticed by SMG and AA and discussed in our weekly meetings.

5.11 Conclusions

As demands on students to practice data presentation increases, it becomes more important for students to practice skills associated with graph choice and construction. In this study we applied evidence-based graphing materials using the CAM into an upper-division, undergraduate inquiry-based physiology lab classroom with aims to advance students towards expertise with graph choice and construction. Despite the limitations of this study, we are confident in methodologies that were used when creating these evidence-based graphing materials, and the instructional strategies that went into implementation of these materials. These graphing materials have the potential to be part of the quantitative literacy conversation, starting at the K-12 levels (NGSS, 2012) and bridging into the undergraduate (AAMC-HHMI Committee, 2009; AAAS, 2011) and graduate biology curricula. By having students practice using these materials early on in elementary school and adjusting the graphing materials according to students' academic and cognitive development levels, we can provide students with appropriate scaffolds and guidance with graphing. This will also allow instructors to diagnose student difficulties with various aspects of graphing as they appear, so that by the time students graduate from college, they have expertise in all components of MRC, allowing them to understand the function of graphs, inventing graphs, critique graphs, and reflect on graphs. Future students should explore the general utility of the graphing materials in a variety of classroom contexts and with students at various levels in their undergraduate education. Additionally, external validation of these graphing materials in biology and non-biology classrooms and with students of different demographics and cognitive levels may reveal novel uses of that may have been overlooked in this classroom study.

Acknowledgements

We would like to thank our research group, PIBERG, for their feedback on this project. We are indebted to all of the biology undergraduate students who participated in this study. This project emerged from ideas initiated within the Biology Scholars Research Residency program.

References

AAMC-HHMI Committee. (2009). Scientific foundations for future physicians. *Washington, DC: Association of American Medical Colleges*.

Ainley, J. (1995). Re-viewing graphing: Traditional and intuitive approaches. For the Learning of Mathematics, 15(2), 10-16.

Anderson, T., & Shattuck, J. (2012). Design-based research a decade of progress in education research?. *Educational researcher*, *41*(1), 16-25.

Angra, A., & Gardner, S. M. (2016). Development of a framework for graph choice and construction. *Advances in physiology education*, *40*(1), 123-128.

Angra, A. & Garnder, S.M. (in review). Evaluating Meta-Representational Competence Along the Expert-Novice Continuum in Biology. *CBE-LSE*.

Bakker, A. (2004). Reasoning about shape as a pattern in variability.*Statistics Education Research Journal*, *3*(2), 64-83.

Barraquand, F., Ezard, T. H., Jørgensen, P. S., Zimmerman, N., Chamberlain, S., Salguero-Gómez, R., ... & Poisot, T. (2014). Lack of quantitative training among early-career ecologists: a survey of the problem and potential solutions. *PeerJ*, *2*, e285.

Bing, T. J., & Redish, E. F. (2012). Epistemic complexity and the journeyman-expert transition. *Physical Review Special Topics-Physics Education Research*, *8*(1), 010105.

Brasell, H. M., & Rowe, M. B. (1993). Graphing skills among high school physics students. School Science and Mathematics, 93(2), 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*(3), 323-332.

Brown, A. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. Journal of the Learning Sciences, 2(2), 141–178.

Collins, A., Brown, J. S., & Holum, A. (1991). Cognitive apprenticeship: Making thinking visible. *American educator*, *15*(3), 6-11.

Collins, A. (1996). Design issues for learning environments. *International perspectives on the design of technology-supported learning environments*, 347-361.

diSessa, A.A., & Sherin, B. L. (2000). Meta-representation: An introduction. *The Journal of Mathematical Behavior*, *19*(4), 385-398.

diSessa, A. A. (2004). Metarepresentation: Native competence and targets for instruction. *Cognition and instruction*, *22*(3), 293-331.

Dennen, V. P., & Burner, K. J. (2008). The cognitive apprenticeship model in educational practice. *Handbook of research on educational communications and technology*, *3*, 425-439.

Gormally, C., Brickman, P., Hallar, B., & Armstrong, N. (2009). Effects of inquiry-based learning on students' science literacy skills and confidence.*International journal for the scholarship of teaching and learning*, *3*(2), 16.

Gormally, C., Brickman, P., & Lutz, M. (2012). Developing a Test of Scientific Literacy Skills (TOSLS): measuring undergraduates' evaluation of scientific information and arguments. *CBE-Life Sciences Education*, *11*(4), 364-377.

Hattikudur, S., Prather, R. W., Asquith, P., Alibali, M. W., Knuth, E. J., & Nathan, M. (2012). Constructing Graphical Representations: Middle Schoolers' Intuitions and Developing Knowledge About Slope and Y-intercept. *School science and mathematics*, *112*(4), 230-240.

Hogan, K., & Maglienti, M. (2001). Comparing the epistemological underpinnings of students' and scientists' reasoning about conclusions.*Journal of Research in Science Teaching*, *38*(6), 663-687.

IBM Corp. Released. (2013). IBM SPSS Statistics for Windows, Version 22.0. Armonk, NY: IBM Corp.

Leonard, J. G., & Patterson, T. F. (2004). Simple computer graphing assignment becomes a lesson in critical thinking. NACTA journal, 48(2), 17-21.

McFarland, J. (2010). Teaching and Assessing Graphing Using Active Learning. *MathAMATYC Educator*, *1*(2), 32-39.

Meletiou-Mavrotheris, M., & Lee, C. (2010). Investigating college-level introductory statistics students' prior knowledge of graphing. *Canadian Journal of Science, Mathematics and Technology Education*, *10*(4), 339-355.

Mevarech, Z. R., & Kramarsky, B. (1997). From verbal descriptions to graphic representations: Stability and change in students' alternative conceptions. *Educational Studies in Mathematics, 32*(3), 229-263.

Ohland, M. W., Loughry, M. L., Carter, R. L., Bullard, L. G., Felder, R. M., Finelli, C. J., …& Schmucker, D. G. (2006, June). The comprehensive assessment of team member effectiveness (CATME): A new peer evaluation instrument. In Proceedings of the 2006 ASEE Annual Conference.

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*(1), 20-26.

Padilla, M. J. (1990). The science process skills. *Research Matters-to the science Teacher, 9004*.

Piantadosi, S. (2013). *Clinical trials: a methodologic perspective*. John Wiley & Sons.

Roth, W. M., & Roychoudhury, A. (1993). The development of science process skills in authentic contexts. *Journal of Research in Science Teaching*, *30*(2), 127-152.

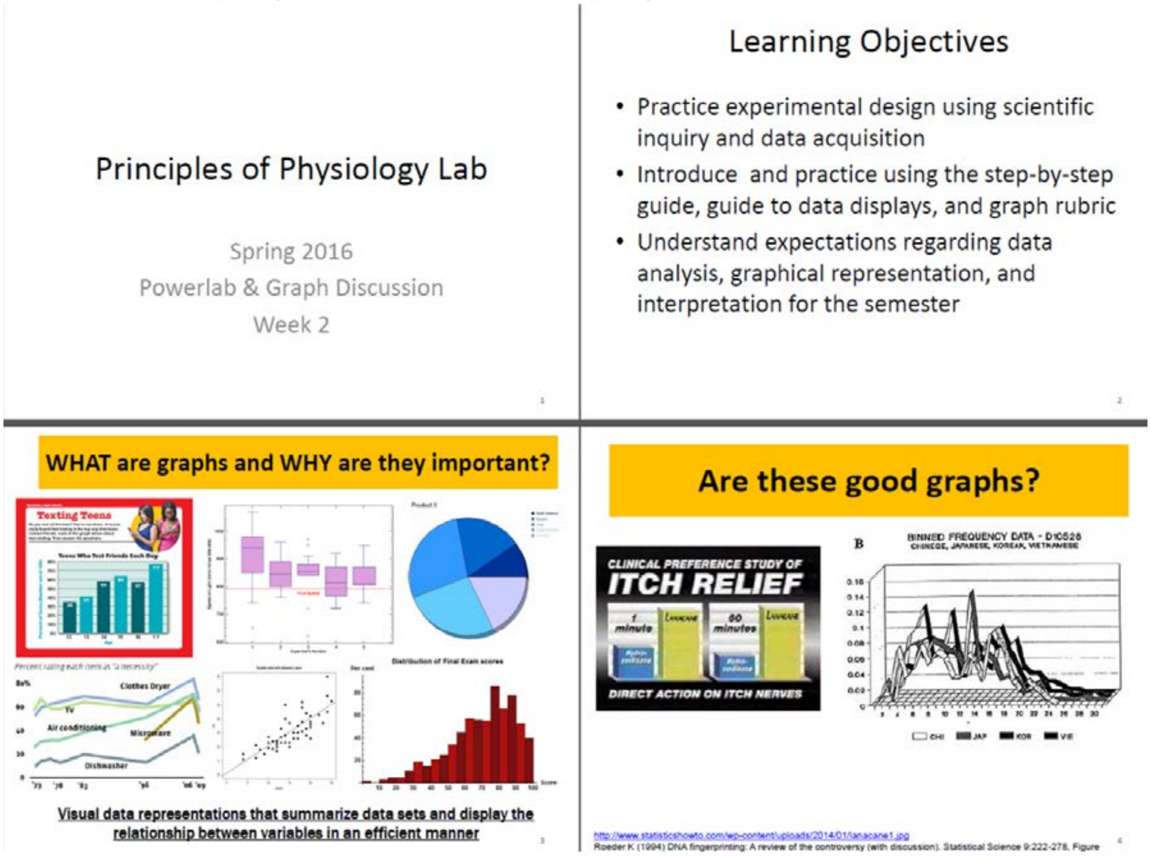
Tairab, H. H., & Khalaf Al-Naqbi, A. K. (2004). How do secondary school science students interpret and construct scientific graphs?. *Journal of Biological Education*, *38*(3), 127-132.

Tynjälä, P. (1999). Towards expert knowledge? A comparison between a constructivist and a traditional learning environment in the university. International journal of educational research, 31(5), 357-442.

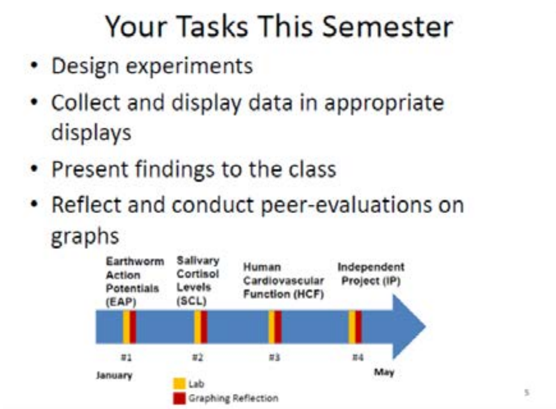
Vygotsky, L.S. (1962). Thought and language. Cambridge, MA: MIT Press (original work published in 1934).

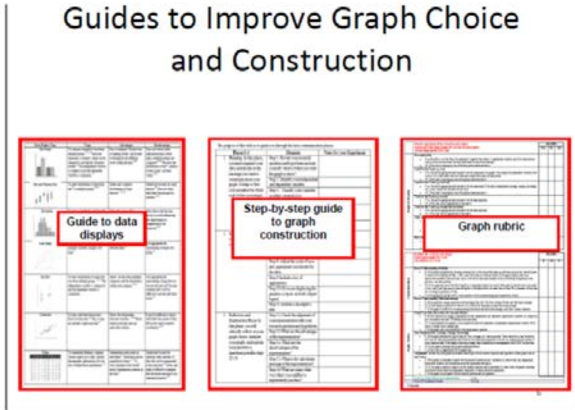
Wolf, K., Stevens, E. (2007) The role of rubrics in advancing and assessing student learning. *The Journal of Effective Teaching, Vol 7, No. 1*, 3-14.

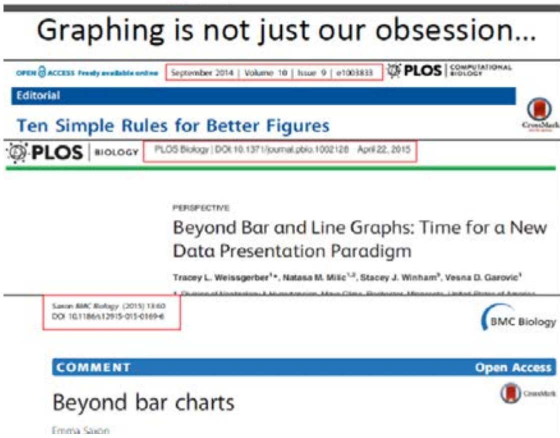
Appendix 5A: PowerPoint Slide from the Mini Lecture on Graphs and Introduction to the Graphing Materials, Week 2 of Spring 2015 and 2016 Semesters

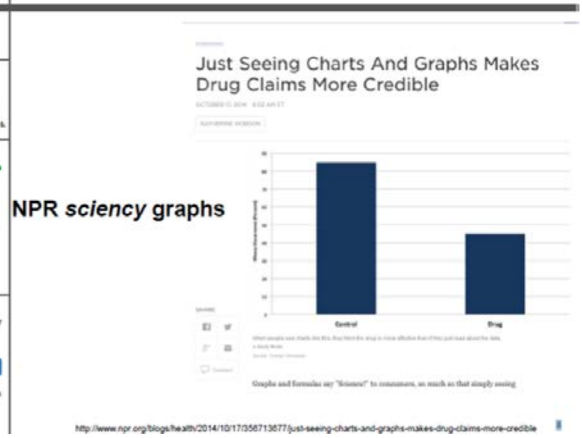


Appendix 5A continued







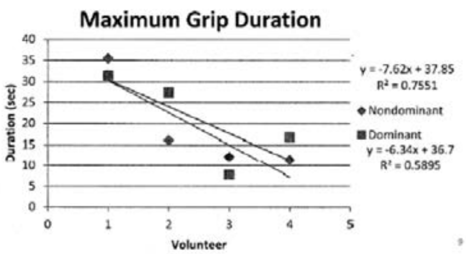


Appendix 5A continued

Question- Is there a difference in the maximum grip duration between the dominant and nondominant hand?

Hypothesis- There is difference in grip duration between the dominant and nondominant hand.

Prediction-Since most of the human population is not ambidextrous, the grip duration for the dominant hand will be significantly higher than the nondominant hand.

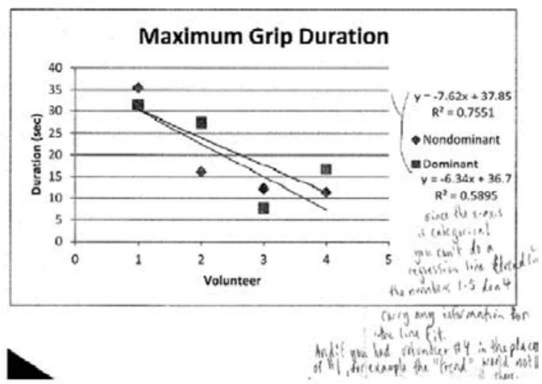


Graph Mechanics

	ID of graph or set of graphs:	Present and Appropriate	Present but Needs Improvement
Graph Mechanics	**Descriptive title** • P.A- Should be: a) in the form of a statement, b) mention the subject, c) appropriate variables, and d) include relevant details about the experiment that help understand the take home message. • NI- If the title is missing any one of the four points mentioned above.		
	Label for the X axis (e.g. time) • P.A- Should be appropriate and descriptive for the experiment. For graphs with categorical independent variables, there needs to be a label under each set of data and a larger label under all data plotted. • NI- If the label is missing any one of the points mentioned above.		
	Label for the Y axis (e.g. heart rate) • P.A- Should be appropriate and descriptive for the experiment. If the data is manipulated (average, change, percentage, etc.), then it should be indicated on the y axis. • NI- If the label is missing any one of the points mentioned above.		
	Units for the X axis (e.g. seconds) • P.A- Should be appropriate and descriptive for the data displayed. • NI- If the units are not appropriate or descriptive.		
	Units for the Y axis (e.g. average beats per minute) • P.A- Should be appropriate and descriptive for the data displayed. • NI- If the units are not appropriate or descriptive.		
	Scale (appropriate intervals and range for data) • P.A- Should be appropriate for the data displayed such that the increments are clear and without clutter and includes appropriate significant figures. If the scale is discontinuous or doesn't start at the origin, it should be indicated by a break in the axis. • NI- If the scale is not appropriate for the data such that it is cluttered, does not include appropriate significant figures, and or if the scale does not indicate axis break.		
	Key (defines different data sets that are plotted) • P.A- Should be appropriate and descriptive for the data displayed. It should include: a) descriptions of different colors (if applicable), b) the sample size and c) the number of trials. • NI- If the key is not descriptive and does not indicate the sample size.		

Graph Communication and Choice

		Excellent (E)	Needs Improvement (NI)	Unsatisfactory (U)
Communication	**Ease of Understanding: Aesthetics** • E- If the graph is aesthetically pleasing, meaning that: a) the data plotted takes up sufficient room in the Cartesian plane, b) makes use of legible size font, c) the x and y axis lines are clear and legible, d) the graph displays data in an appropriate number of bars and lines, and e) is devoid of chart junk elements such as: distracting background colors, patterns, and dark gridlines. • NI- If the graph has one of the following flaws: a) the graph displays too much white space, b) the font size is too small, c) the x and y axis lines are not clear and legible, d) the graph shows too many bars or lines OR e) elements of chart junk are clouding interpretation of data. • U- If the graph has multiple flaws, which interfere with the understanding and interpretation of data.			
	Ease of Understanding: Take home message • E- If the graph has sound construction and mechanics that allow for clear sorting of trends and take home message. • NI- If data trends are difficult to observe or it is difficult to formulate a proper take home message. • U- If the graph is ineffective at communicating data trends and take home message, such that it causes confusion.			
Graph Choice	**Graph Type (Bar, line, scatter, dot, box and whisker)** • E- If data displayed in a graph is appropriate for both independent and dependent experimental variables (i.e. categorical and continuous) and data. (*Referring to the data form) • NI- If data displayed in a graph is a) not suitable for either the dependent or independent experimental variables OR b) there is a better way to present data. • U- If the graph type is not suitable for both experimental variables.			
	Data Displayed (Raw, Averages, Changes, Percentage) • E- If the graph indicates the type of data (ex. Raw, averages, etc.) that are plotted. There should be a clear distinction between raw data and manipulated data based on the information presented in the key (ie. sample size and number of trials) and axis label. If the graph is showing averages, then it should also be accompanied with STDEV or error bars. • NI- If the graph is missing one of points mentioned above. • U- If data type is inappropriate for the graph type			
	Alignment *(at least one of the graphs presented should align with the research question and hypothesis. Other graphs can be exploratory.)* • E- If the graph is completely aligned with the research question and or hypothesis. In other words, the independent, dependent variables, and information about the experiment are explicit. • NI- If the graph is partially aligned with the research question and or hypothesis. In other words, the graph is missing information about either the independent, dependent, or details about the experiment. • U- If the graph is not aligned with the research question and or hypothesis.			



Appendix 5A continued

Reflecting on a Graph

A good reflection should address the advantages and the disadvantages.

- Helpful questions to ask yourself as you reflect:
  - **Is the graph communicating a clear take-home message?**
  - **Are the experimental variables (categorical, continuous, independent, dependent) displayed, appropriate for this graph type?**
  - **Is the graph displaying data that are suitable for this graph type?**
  - **Is the graph aligned with the research question and hypothesis?**
- **If the answer is "yes" to any of these, this is an advantage of the graph.**
- **If the answer is "no" to any of these, this is a disadvantage and you should provide suggestions for improvement!**

13

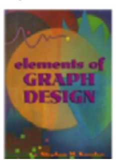
Example

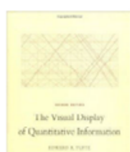
- **Advantage**- The scatter plot accurately displays the dependent variable on the y axis with appropriate units. The data displayed on the graph appear to be raw data, which is appropriate for a scatter plot.
- **Disadvantage**- The scatter plot does not align with the research question and hypothesis. The hypothesis states finding a difference between the dependent variable yet the graph type and data displayed are not comparative. In order to align these data with the hypothesis, I suggest taking the grip duration difference between the dominant and nondominant hands for each trial, for each subject, and calculate an average difference in duration, calculate the standard error, and display the data in a bar graph. This way, the message the graph is trying to communicate will be clear and effective.

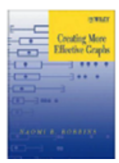
14

Useful Resources on Good and Bad Graphs

- Kosslyn, S. M. (1994). *Elements of graph design*. New York: WH Freeman.
- Tufte E (2001) *The visual display of quantitative information*, 2nd edition. Graphics Press.
- Robbins NB (2004) *Creating more effective graphs*. Wiley.







Appendix 5B: Point Values Associated with the Graph Rubric for Research Purposes

Research Question:

Hypothesis:

	Present/Appropriate (P/A)= 0.5 pts for each category Present but Needs Improvement (NI) =0.25 pts for each category Absent/Inappropriate (X/I)=0 pts	GRAPH 1			GRAPH 2		
		P/A	NI	X/I	P	NI	X/I
Graph Mechanics	**Descriptive title** • P/A-Should be: a) in the form of a statement, b) mention the subject, c) appropriate variables, and d) include relevant details about the experiment that help understand the take home message. • NI- If the title is missing any one of the four points mentioned above.						
	Label for the X axis (e.g. time) • P/A- Should be appropriate and descriptive for the experiment. For graphs with categorical independent variables, there needs to be a label under each set of data and a larger label under all data plotted. • NI- If the label is missing any one of the points mentioned above.						
	Label for the Y axis (e.g. heart rate) • P/A- Should be appropriate and descriptive for the experiment. If the data is manipulated (average, change, percentage, etc.), then it should be indicated on the y axis. • NI- If the label is missing any one of the points mentioned above.						
	Units for the X axis (e.g. seconds) • P/A- Should be appropriate and descriptive for the data displayed. • NI- If the units are not appropriate or descriptive.						
	Units for the Y axis (e.g. average beats per minute) • P/A- Should be appropriate and descriptive for the data displayed. • NI- If the units are not appropriate or descriptive.						
	Scale (appropriate intervals and range for data) • P/A- Should be appropriate for the data displayed such that the increments are clear and without clutter and includes appropriate significant figures. If the scale is discontinuous or doesn't start at the origin, it should be indicated by a break in the axis. • NI- If the scale is not appropriate for the data such that it is cluttered, does not include appropriate significant figures, and/or if the scale does not indicate axis break.						
	Key (defines different data sets that are plotted) • P/A- Should be appropriate and descriptive for the data displayed. It should include: a) descriptions of different colors (if applicable), b) the sample size and c) the number of trials. • NI- If the key is not descriptive and does not indicate the sample size.						
	Total Points for Mechanics: /3.5pts						

Appendix 5B continued

	Excellent (E) = 2 pts for each category Needs Improvement (NI) = 1 pts Unsatisfactory (U) = 0 pts	GRAPH 1			GRAPH 2			GRAPH 3		
		E	NI	U	E	NI	U	E	NI	U
Communication	**Ease of Understanding-Aesthetics** • E- If the graph is aesthetically pleasing, meaning that: a) the data plotted takes up sufficient room in the Cartesian plane, b) makes use of legible size font, c) the x and y axis lines are clear and legible, d) the graph displays data in an appropriate number of bars and lines, and e) is devoid of chart junk elements such as: distracting background colors, patterns, and dark gridlines • NI- If the graph has one of the following flaws: a) the graph displays too much white space, b) the font size is too small, c) the x and y axis lines are not clear and legible, d) the graph shows too many bars or lines OR e) elements of chart junk are clouding interpretation of data. • U- If the graph has multiple flaws, which interfere with the understanding and interpretation of data.									
	Ease of Understanding-Take home message • E- If the graph has sound construction and mechanics that allow for clear sorting of trends and take home message. • NI- If data trends are difficult to observe or it is difficult to formulate a proper take home message. • U- If the graph is ineffective at communicating data trends and take home message, such that it causes confusion.									
Graph Choice	**Graph Type (Bar, line, scatter, dot, box and whisker)** • E- If data displayed in a graph is appropriate for both independent and dependent experimental variables (i.e. categorical and continuous) and data. (*Referring to the data form) • NI- If data displayed in a graph is a) not suitable for either the dependent or independent experimental variables OR b) there is a better way to present data. • U- If the graph type is not suitable for both experimental variables.									
	Data Displayed (Raw, Averages, Changes, Percentage) • E- If the graph indicates the type of data (ex. Raw, averages, etc.) that are plotted. There should be a clear distinction between raw data and manipulated data based on the information presented in the key (ie. sample size and number of trials) and axis label. If the graph is showing averages, then it should also be accompanied with STDEV or error bars. • NI- If the graph is missing one of points mentioned above. • U- If data type is inappropriate for the graph type									
	Alignment *(at least one of the graphs presented should align with the research question and hypothesis. Other graphs can be exploratory.)* • E- If the graph is completely aligned with the research question and/or hypothesis. In other words, the independent, dependent variables, and information about the experiment are explicit. • NI- If the graph is partially aligned with the research question and/or hypothesis. In other words, the graph is missing information about either the independent, dependent, or details about the experiment. • U- If the graph is not aligned with the research question and/or hypothesis.									
	Total for Graph Choice and Communication: /10pts									
	Overall Presentation Grade: /13.5 pts									

Appendix 5C: Example of Feedback Given using the Graph Presentation Rubric

Research Question: Are the medial and lateral axons of an earthworm uniform throughout the worm?

Hypothesis: Because earthworms are annelids, their medial and lateral axons are largely uniform throughout the length of a worm.

	E = 1 pt for each category NI = 0.5 pts U = 0 pts	Excellent (E)	Needs Improvement (NI)	Unsatisfactory (U)
Presentation	**Introduction (Background information/rationale, hypothesis, prediction)** • E, if an appropriate background and rationale are stated that naturally leads to the QHP. Hypothesis and prediction are formulated properly (e.g. hypothesis should be a testable and falsifiable statement of how you think things work and the prediction is based on the hypothesis). • NI, if the background and rationale are somewhat related to QHP and/or the hypothesis and predictions are not completely formulated properly. • U, if the background and rationale are not developed and/or are unrelated to QHP and the hypothesis and predictions are not formulated properly.	Interesting worm anatomy	Elaborate more on why axon diameter matters	
	Basic experimental approach • E if the methods are clear and state what was varied or manipulated. Controls or comparison groups are defined. • NI if it is not clear exactly how the experiment was performed or controls and comparison groups are not clearly defined. • U if it is not clear how the experiment was performed and controls and comparison groups are not clearly defined.	good raw data to illustrate measurement good segmented and assume uniformity		
	Description of the data • E if the presenter walks the audience through each data representation, defining what is plotted, noting trends, and stating the take-home message • NI if it the presenter moves too quickly through the data without defining what is plotted or stating the trends and take-home message. • U if the presenter does not present the data.		Walk us through the axes Describe the data points	
	Conclusions (take-home message, hypothesis, limitations, future directions) • E if the presenter walks states the take-home message from the study, how the findings relate to the hypothesis and states limitations and future directions for the project. • NI if it the presenter superficially moves through the take-home message or does not reflect on the hypothesis, limitations, or future directions. • U if the presenter omits any 3 of the 4 components	Good ideas on limitations and for future directions		
Oral Communication	**Clarity** • E if the presentation is easy to follow and it is clear that the talking points were planned and each person knows what they want to say. • NI if parts of the presentation are hard to follow. • U if it is difficult to follow the entire presentation.	x		
	Total for Presentation and Oral Communication: 4 /5 pts			

Appendix 5C continued

	Present /Appropriate (P/A)= 0.25 pts for each category Present but Needs Improvement (NI) =0.125 pts for each category Absent/Inappropriate (X/I)= 0 pts	GRAPH 1			GRAPH 2		
		P/A	NI	X/I	P	NI	X/I
Graph Mechanics	**Descriptive title** • P/A-Should be: a) in the form of a statement, b) mention the subject, c) appropriate variables, and d) include relevant details about the experiment that help understand the take home message. • NI- If the title is missing any one of the four points mentioned above.		Doesn't reflect what you were looking at				
	Label for the X axis (e.g. time) • P/A- Should be appropriate and descriptive for the experiment. For graphs with categorical independent variables, there needs to be a label under each set of data and a larger label under all data plotted. • NI- If the label is missing any one of the points mentioned above.	x					
	Label for the Y axis (e.g. heart rate) • P/A- Should be appropriate and descriptive for the experiment. If the data is manipulated (average, change, percentage, etc.), then it should be indicated on the y axis. • NI- If the label is missing any one of the points mentioned above.		Hard to relate this to your experiment	And your Q and H are a little vague			
	Units for the X axis (e.g. seconds) • P/A- Should be appropriate and descriptive for the data displayed. • NI- If the units are not appropriate or descriptive.	X					
	Units for the Y axis (e.g. average beats per minute) • P/A- Should be appropriate and descriptive for the data displayed. • NI- If the units are not appropriate or descriptive.	X;					
	Scale (appropriate intervals and range for data) • P/A- Should be appropriate for the data displayed with appropriate significant figures. If the scale is discontinuous or doesn't start at the origin, it should be indicated by a break in the axis. • NI- If the scale is not appropriate for the data and/or if the scale does not indicate axis break.		X; start at the origin or have an axis break				
	Key (defines different data sets that are plotted) • P/A- Should be appropriate and descriptive for the data displayed. It should include: a) descriptions of different colors (if applicable), b) the sample size and c) the number of trials. • NI- If the key is not descriptive and does not indicate the sample size.	n/a					
	Total Points for Mechanics: 1.5 /1.75pts						

Appendix 5C continued

	Excellent (E) = 0.65 pts for each category Needs Improvement (NI) = 0.325 pts Unsatisfactory (U) = 0 pts	GRAPH 1			GRAPH 2		
		E	NI	U	E	NI	U
Communication	**Ease of Understanding-Aesthetics** • E- If the graph is aesthetically pleasing, meaning that: a) the data plotted takes up sufficient room in the Cartesian plane, b) makes use of legible size font, c) the x and y axis lines are clear and legible, d) the graph displays data in an appropriate number of bars and lines, and e) is devoid of chart junk elements such as: distracting background colors, patterns, and dark gridlines • NI- If the graph has one of the following flaws: a) the graph displays too much white space, b) the font size is too small, c) the x and y axis lines are not clear and legible, d) the graph shows too many bars or lines OR e) elements of chart junk are clouding interpretation of data. • U- If the graph has multiple flaws, which interfere with the understanding and interpretation of data.		Don't need grid lines				
	Ease of Understanding-Take home message • E- If the graph is constructed in a way that is: a) clear to sort trends and b) easy to note the take home message. • NI- If data trends are difficult to observe or it is difficult to formulate a proper take home message. • U- If the graph is ineffective at communicating data trends and take home message.		Hard to understand what you are plotting; why not conduction velocity?				
Graph Choice	**Graph Type (Bar, line, scatter, dot, box and whisker)** • E- If data displayed in a graph is appropriate for both independent and dependent experimental variables (i.e. categorical and continuous) and data. (*Referring to the data form) • NI- If data displayed in a graph is a) not suitable for either the dependent or independent experimental variables OR b) there is a better way to present data • U- If the graph type is not suitable for both experimental variables.	x					
	Data Displayed (Raw, Averages, Changes, Percentage) • E- If the graph indicates the type of data (ex. Raw, averages, etc.) that are plotted. There should be a clear distinction between raw data and manipulated data based on the information presented in the key (ie. sample size and number of trials) and axis label. If the graph is showing averages, then it should also be accompanied with STDEV or error bars. • NI- If the graph is missing one of points mentioned above. • U- If data type is inappropriate for the graph type	X; raw probabl y					
	Alignment *(at least one of the graphs presented should align with the research question and hypothesis. Other graphs can be exploratory.)* • E- If the graph is completely aligned with the research question and/or hypothesis. In other words, the independent, dependent variables, and information about the experiment are explicit. • NI- If the graph is partially aligned with the research question and/or hypothesis. In other words, the graph is missing information about either the independent, dependent, or details about the experiment • U- If the graph is not aligned with the research question and/or hypothesis.		x				
	Total Points for Graph Choice and Communication: 2.275 /3.25pts						
	Overall Graph Presentation Grade: 7.775 /10 pts						

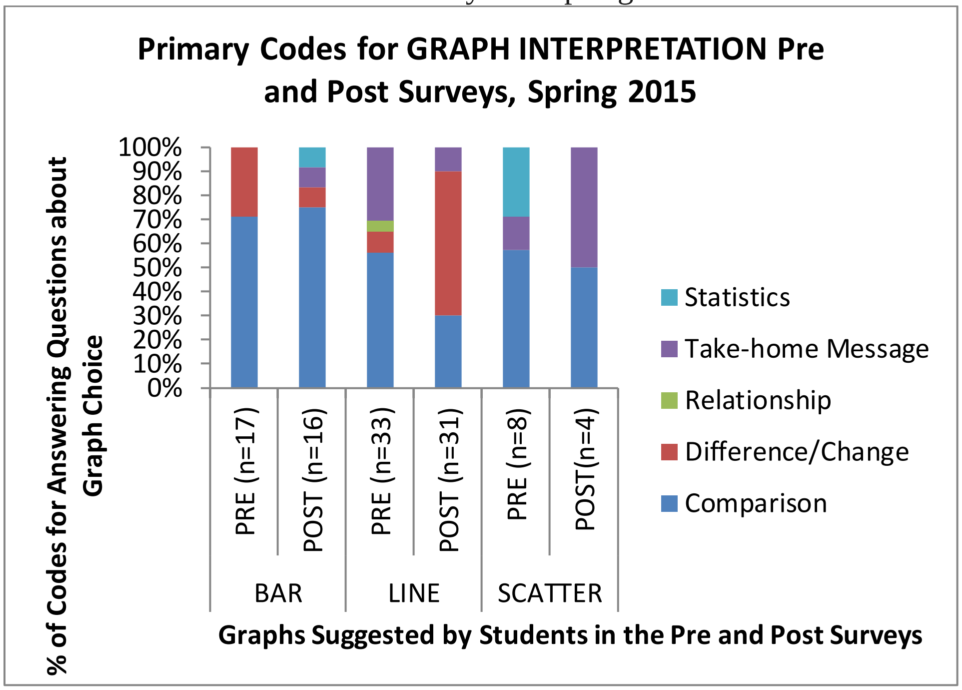
Appendix 5D: Questions Used in the Pre-survey for the Intervention Semesters.

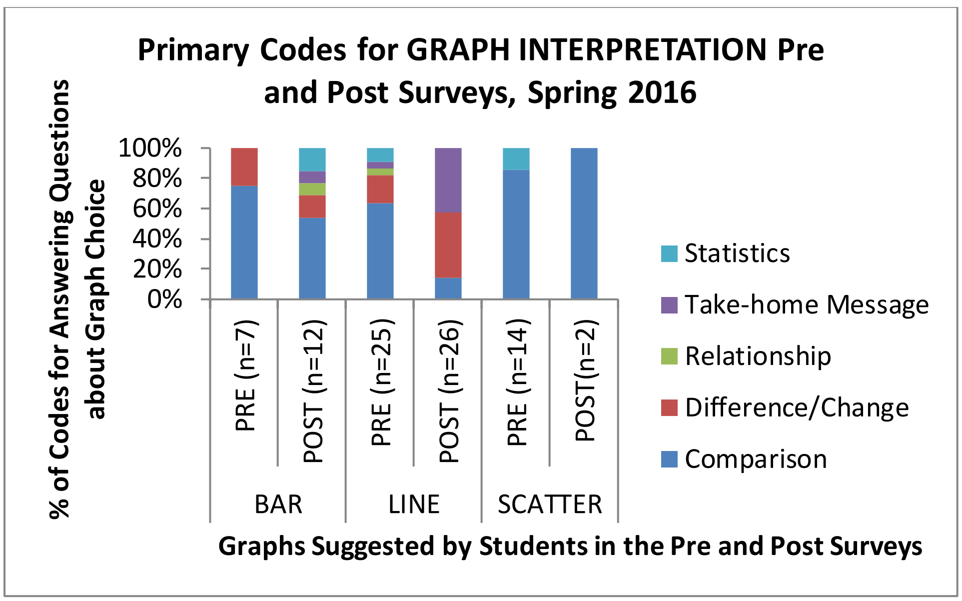
What is your name?
Which of the following best describes your class standing?
Which of the following best describes your class standing?
Which of the following describes your undergraduate study track at Purdue?
Which of the following describes your undergraduate study track at Purdue?
Have you thought about what you would like to do upon the completion of your undergraduate degree?
Have you taken statistics at Purdue University? If yes, please list the course number.
Do you have any research experience? This can consist of high school science fair projects, undergraduate project-based labs, etc.
If you selected 'Yes' for the question above, please briefly explain the type of data you worked with.
What is the most common method that you use to analyze data and construct graphs?
How would you define a graph?
What are graphs used for?
What are parts of a graph? List as many as you can think of.
Please use the following scenario to answer the next set of questions to the best of your ability
Formulate a research question based on the plant scenario given above.
Formulate a hypothesis for the plant scenario above.
What type of graph would you make for the data provided in the table above?
Why did you choose this option over other graph types?
Please describe what you would plot on the x-axis? Why?
Please describe what you would plot on the y-axis? Why?
When did you learn about this type of graph?

Appendix 5E: Questions Used in the Post-survey for the Intervention Semesters.

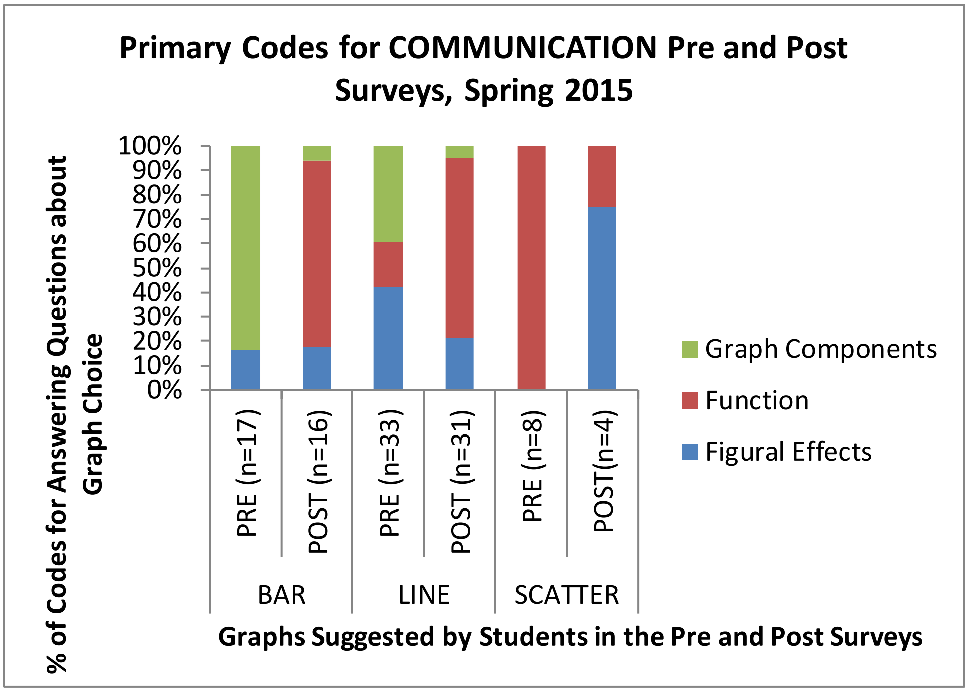
What is your name?
Please use the following scenario to answer the next set of questions to the best of your ability
Formulate a research question based on the plant scenario given above.
Formulate a hypothesis for the plant scenario above.
What type of graph would you make for the data provided in the table above?
Why did you choose this option over other graph types?
Please describe what you would plot on the x-axis? Why?
Please describe what you would plot on the y-axis? Why?
When did you learn about this type of graph?
Please rate the likelihood of utilizing each learning material in your future classes.
Overall, how confident do you feel with graph construction?
Overall, how confident do you feel with evaluating graphs?
Is there anything else you would like to tell us about your learning experiences with graphing this semester?
Do you have any comments and/or suggestions for improving the learning materials?

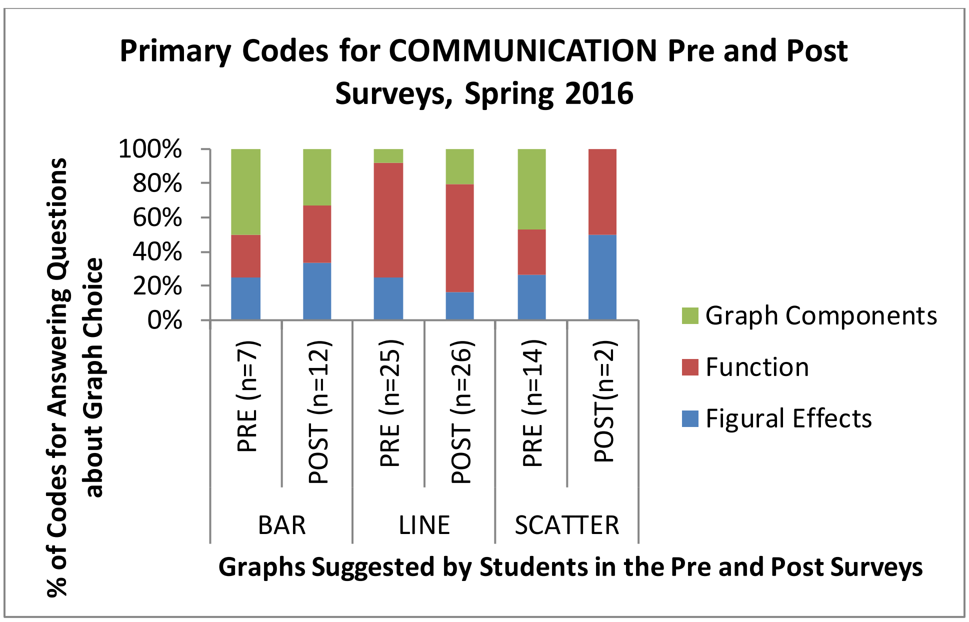
Appendix 5F: Primary Codes for Graph Interpretation and Communication for Pre and Post Surveys for Spring 2015 and 2016





Appendix 5F continued



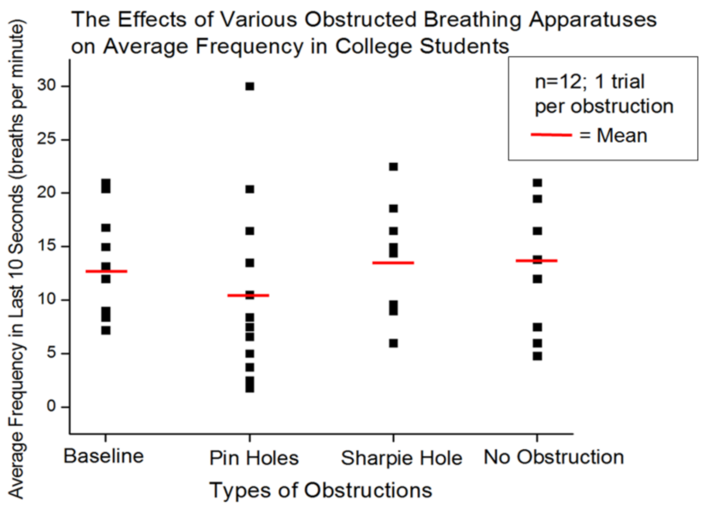


Appendix 5G: Examples of Graph Alignment with the Research Question and Hypothesis

Criteria for Graph Alignment, taken from the Graph Rubric
Alignment (at least one of the graphs presented should align with the research question and hypothesis. Other graphs can be exploratory.)
E- If the graph is completely aligned with the research question and/or hypothesis. In other words, the independent, dependent variables, and information about the experiment are explicit.
NI- If the graph is partially aligned with the research question and/or hypothesis. In other words, the graph is missing information about either the independent, dependent variable, or details about the experiment.
U- If the graph is not aligned with the research question and/or hypothesis.

Example of Complete Alignment

Research Question- How do varying airway obstructions affect the average final tidal volume, frequency, and minute ventilation of college students?
Hypothesis- Obstructed air flow will have an effect on tidal volume, frequency and minute ventilation, but varying obstructions will have different effects.

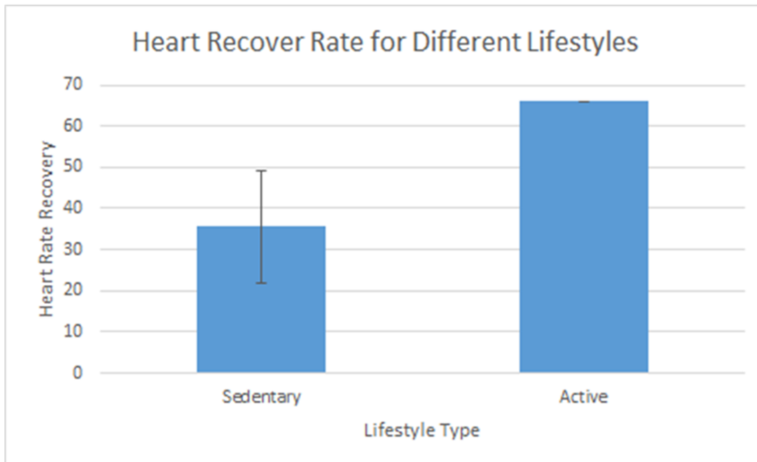


AA and SMG's comment: "*This graph is completely aligned with the research question and hypothesis because the independent, dependent variables, and information about the experiment are explicit.*"

Appendix 5G continued

Example of Partial Alignment

Research Question- Does lifestyle affect heart rate recovery?
Hypothesis- Lifestyle does affect heart rate recovery.

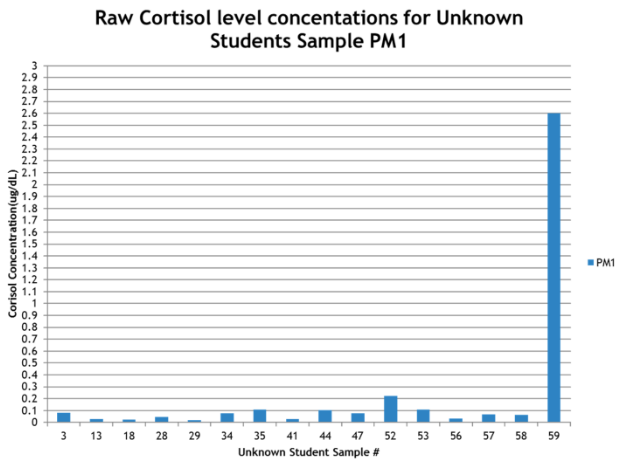


AA and SMG's comment on alignment: *"The graph is partially aligned because information on the experiment is not explicit. Were you looking at exercise?"*

Appendix 5G continued

Example of No Alignment

Research Question: Did the academic stress associated with the Biology 328 exam have an effect on students' PM 1 and AM1 cortisol concentrations?
Hypothesis: The academic stress associated with the Biology 328 exam will result in abnormal variations in salivary cortisol concentrations for the PM1 and AM1 sampling periods.



AA and SMG's comment on alignment: *"This graph is not aligned with the research question and hypothesis because it is unclear that the graph is showing cortisol concentration associated with academic stress of the biology exam. The graph is not displaying data for the AM1 data point and the type of cortisol is not specified."*

Appendix 5H: Students' Feedback on the Graphing Materials from the Intervention Semesters

I feel like I learned a lot this semester about graphs. The rubrics helped out a lot.

I definitely think I am more confident in my graphing abilities because of this semester and am more detailed when making them.

I gained a lot more knowledge about how to construct the graphs correctly with all the information needed.

I certainly learned how important it is to incorporate certain elements into a graph! I cannot believe how much I was missing before.

I have gained a better understanding of what situations it is appropriate to use different types of graphs.

You want your graph to be relative to the research question and contain all the information needed to understand it on it without much background explanation

I believe many people who were unfamiliar with graphs improved greatly this semester.

I learned a lot of new material involving graphing.

Graphing feedback was very useful the first few weeks of lab!

I thought this classed allowed me to develop oral communication skills, organizing presentations, and analyzing data

I learned a lot . I enjoyed putting our own experiments in graphical representations. It was a fun way to learn. The application part of it was good.

The step by step guide is very helpful

I have not done this many presentation before and this definitely helped me to improve giving presentation

I learned more about how to make graphs more visually pleasing and how to make them easier for others to understand.

I learned that there is more to graphing than just plotting data

I feel like I learned how to make better graphs that more clearly show my data

I feel that my graphing skills (as well as interpreting other graphs) have improved with doing so many experiments.

It took me awhile to understand all the necessary components in graphs. This class helped me understand when certain graph types are relevant and also what to include on each graph type.

Peer graph evaluations were helpful! Particularly in determining how easy it is to understand the take home message and whether it matches the hypothesis/predictions.

I'm confused when to used average data vs. raw data.

I learned a lot, but it still seemed like all we ever did were line or bar graphs. I had hoped to get more experience with more "grown up" ways of presenting data.

seeing a number of graphs and evaluating was helpful, but i didn't feel like receiving the feedback from peers was that helpful

Appendix 5H continued

The peer evaluations received from other groups weren't as helpful as I thought. I usually only glanced at them.
I believe I have improved with graph formation, but still am not completely comfortable in knowing which to use right off the bat.
Thank you for taking the time out to actually make sure that we understand the material.
I thought the peer review was most helpful
Please indicate the parts that need to be included in each type of graphs.
Provide more feedback on the level of detail you are looking for in graphs eg. titles, etc.
I enjoyed this class and I feel much more confident in my ability to build graphs
Thanks for introducing me to statistical analysis!
I learned a lot, but still having trouble with two-tailed t test graphs
It was really good experience working with my group members and learning how to construct graphs and evaluate them.
I have learned a lot about how to manipulate graphs effectively in order to present information to my classmates
That it will be applicable to future projects
I feel that I didn't learn about a wide variety if graphs, mostly focused on a few of them.
It was helpful and challenging to have to present our data every week in the most effective way possible, thank you for all the helpful graphing tools!
I feel like the grading of them could have been a little more consistent. I noticed the last couple of weeks that the graders were getting more particular. I understand this reflects our growth, but I feel like we should have gotten marked off for things consistently and then be expected to incorporate them later. That would be a more accurate reflection of our growth.
Thanks for a great semester! I learned a lot of information especially in the area of graph construction!
When evaluating graphs, it would be better to have more than 2 or 3 categories. Sometimes people's graphs were not perfect, but I felt like they deserved more than half credit or no credit for that particular category
I have broadened my knowledge of graphs
I knew how to use graphs but confidence is much higher now. I have learned to make sure the data presented matches the hypothesis.
no! I enjoyed the approach to teaching us about forming graphs from the data we collected
I liked the systematic way the rubric made us look at graphs.
After number of our experience with various type of graph, I was able to better understand correlation between data points and also the bigger picture on our lab assignments

Appendix 5H continued

It's just plotting graphs on computer software that I have a slight problem with, but manually, I'm confident with it,
Went well, and activities helped
I think the handouts at the beginning of the semester were very useful as well as the graph feedback from each presentation. I know my graph looked at those before creating the graph for each experiment so that we were constantly improving and getting as many points as possible.
I liked the way the class was set up
was certainly great getting formal repetitive work and evaluation of graphs
Aakanksha showed us a really interesting dot-plot graph that was like a box-plot, showing each data point for each group in a categorical variable and having the mean shown with a bar. I liked that graph and it would be awesome if we could find a way to make one of those in this class!
Spend a little bit more time on explaining all the types of graphs and when they would be most useful to use
Origin is very helpful once you understand how to use it! But the current system is very effective and I would not change much if anything.
Information on which graphs match which types of data sets is helpful
I wish the experiments we did were more applicable to different types of graphs. I feel that we used the same graphs over and over because of the types of experiments we performed.
I'm not the biggest fan of grading peer's graphs because I feel like I didn't even look at the peer evaluations.
Maybe having a day where we go over statistical analysis since a lot of us since haven't taken a statistics course.
The peer feedback was not very useful. I learned more from TA and professor feedback.
Make groups use all forms of graphs.
In addition to spitting the groups up it would be beneficial for the groups to distribute work load in a way that encourages everyone to be involved in graph formation, not just those individuals that are comfortable with producing them.
When groups are setting up their graphs the week before a presentation, help guide them in the right direction. A lot of times we were getting ready to present and someone would come tell us to change our graph. By then it was too late.
INCLUDE A STATISTICS GUIDE. No one in my group, including myself, had ever taken a statistics course. I felt very lost and useless in terms of doing that portion of our projects. I would go so far as to say statistics should be required as a prerequisite for this course.
Physical copies f the graph guide for every student may be more helpful than a single copy for each group.

Appendix 5H continued

I wish we would have had a lab session where we went into more detail about making certain graphs and making statistical analysis like one-tailed and two-tailed t-test, and anova

The peer evals of the graph seemed like busy work. I prefer a teacher’s feedback rather than my peers.

I think it is helpful to talk through your graphs with the TA in person rather than reading the TA's feedback. This way, you can ask questions about the evaluation of your graph.

We didn't get a chance to work with a lot of different types of data, it was mostly bar or line graphs

give more mini lectures on graphs

toward the beginning of the lab semester, it would be helpful top be reminded of the types of graphs and what kind of data they are useful for. Sometimes it is hard to remember when for most lab presentations we use the same types over and over. We never have the flexibility or power to decide what other graphical options we could use for our data sets. Most professors are set in their ways and tell you what to use.

All of the materials were really useful and helped a lot. I can't think of any improvements.

I am satisfied with the learning materials.

Some suggestions for our graphs were not given to us until the last labs, so we had been not including that info all semester (human study). It would have been nice to have received sooner.

Maybe more specific rubrics. Sometimes I felt that we were missing points on "new" things every week even though we had been doing them consistently until then.

The peer reviews were not as helpful as the other learning tools

It has been two years since I took a stat class and a little refresher in the beginning of the semester would have helped

More presentation would be ideal

Be more open minded to how groups would like to present their materials not just the way the TA or teacher would want it

Be more consistent week to week about what is needed to put in the graphs, sometimes it would be unclear about what we needed to include to make it the best graphical representation of our data possible.

I really like peer editing graphs and giving feedback. The only thing that I did not like was sometimes it felt like people were just looking for things to comment on so they would write things that didn't apply to our graph or marking off for things that were present. It would be helpful to discuss with the groups that evaluated us to get a better understanding of where our graphs lacked.

I did not like the peer review because I felt that they usually differed greatly from what the instructors evaluated us.

Appendix 5H continued

The materials talked about all different kinds of graphs, but we really only got the chance to make bar graphs and a couple line graphs. The other types were not appropriate for our experiments, but maybe doing experiments that use pie graphs or whatever would be good practice.
Define better what a hypothesis is. I am still extremely lost on how to write a proper hypothesis.
Maybe a take home assignment where we create a graph on our own without our groups?
Including figure descriptions with peer graph evaluations.
Maybe some video clip links for better understanding of the materials would be useful.

CHAPTER 6: CONCLUSIONS

6.1 Dissertation Focus

Graphs condense large amounts of data and display the relationship between variables in an organized visual manner. Graphical displays are plentiful in print and electronic media where they are used to inform the general public of news and events (Monteiro & Ainley, 2007) Graphs are especially ubiquitous to the sciences, including the area of biology, where they are used to assist in the communication of findings from experiments. Creating and interpreting graphical representations are important components of science literacy and involve quantitative and disciplinary knowledge (Picone et al., 2007; AAAS, 2011; Gormally, Brickman, and Lutz, 2012) and are necessary for preparing undergraduate students for medical school (AAMC/HHMI, 2009) and graduate school (Barraquand et al., 2014). Given the importance of graphs, however, not many studies have been conducted at the undergraduate level and graphing materials aimed at improving reasoning with graphing do not exist. Operating under the constructivist paradigm (Duffy & Jonassen, 1992; Bodner, 2006; Bodner & Orgill, 2007), we addressed these gaps in the graphing literature, by first conducting two studies to understand how experts and novice biologists create graphs in a clinical setting (chapter 2), and how students create graphs in a naturalistic classroom environment, using their own data for graph construction (chapter 3). Findings from these two studies were compiled with previously reported student difficulties at the K-12 and undergraduate levels, and informed the development of evidence-based graphing materials for instructors and students to target previously established graphing difficulties (chapter 4; Angra & Gardner, 2016). Lastly, the impact of these evidence-based graphing materials embedded within the cognitive apprenticeship model (Collins et al., 1991) of instruction was tested in an inquiry-based upper-level physiology lab over two semesters (chapter 5).

6.2 Contributions to the Graphing Literature in Undergraduate Biology Learning

As discussed below, this dissertation adds to available graphing literature in 4 ways.

Finding #1: Graphing instructional materials, expert feedback, and repeated practice improve learning about graph choice and construction

The first general finding is that graphing is a skill that cannot be perfected on its own, even for students at the journeyman level who are engaged in undergraduate research in private laboratories (chapter 2) or those enrolled in inquiry-based laboratories (chapters 3 and 5). Just like any skill, the novice student requires: appropriate scaffolds, targeted and frequent feedback from an instructor or mentor who has achieved expertise, and novel contexts to apply their newly acquired ideas (Collins et al., 1989; Roth & Bowen 2001; McFarland 2010). In chapter 5, we demonstrated this process with undergraduate students enrolled in an upper-division physiology lab course. These students received instruction on the usage of the step-by-step guide, guide to data displays, and the graph rubric to help them with their graphing. Students were given expert feedback on their graph mechanics, communication, and choice after each graph presentation. Findings (chapter 5) revealed that it took four iterations with targeted feedback to notice highly significant differences (Kruskal Wallis ANOVA, $p<.0005$) in students' graph construction and across all three graph rubric categories: mechanics, communication, and choice.

Finding #2: Students did not reflect on the research question and/or hypothesis when reflecting on their graph choice.

Alignment of the graph with the research question and/or hypothesis is a subcategory of graph choice in the graph rubric. Despite our efforts in the intervention semester (chapter 5) (e.g. having students generate a research question and hypothesis for every experiment for their own data, generated in an inquiry-based lab, stating the research question and hypothesis for their graph critiques, giving feedback on their graph with the rubric, and making it a part of the step-by-step guide), students did not align their graph choice in the post-graph survey to their research question and/or hypothesis. It is important to develop graphs in the context of particular hypotheses and research questions, because this allows that graph to communicate a specific message. Across all graph studies (chapter 2, 3,5), we noticed the same problem of lack of alignment with the research question and hypothesis. In the plant and bacteria growth scenarios used in chapter 2, we did not explicitly provide the research question and hypothesis to the participants but provided a description of the task, to see if participants would a) identify the purpose of the task, b) construct a graph that aligns with the task and c) articulate their graph choice reflection with the purpose mentioned in the graphing task. Data from the clinical think-aloud interviews revealed that almost all undergraduate students did not identify the purpose of the task nor did they articulate their graph choice to the purpose of the graphing task. However, we did see two instances with professors and two instances with graduate students, where they reflected on the purpose of the task their graph was fulfilling when reflecting on their graph choice. In future graphing interviews, participants should be asked explicitly to state a research question and hypothesis before proceeding with graph construction to see if there is alignment with their graph. The interviewer can also probe the participant to reflect on the research question and hypothesis as they reflect on their graph, and to explain why they did this. Future instruction should explicitly incorporate research questions and hypotheses into graph alignment. One way to achieve this in a classroom setting is by having instructors use the graph rubric (as we did in Chapter 5) to model graph alignment with graph choice and encouraging students to practice aligning their graphs to their research question and hypothesis. Furthermore, allowing students opportunities to critique graphs in textbooks and journal articles (see McFarland, 2010) will increase students' MRC, and may help students transition into becoming critical thinkers and being better evaluators of data and evidence.

Finding #3: Undergraduate students and expert professors created line graphs when given the bacteria or plant scenarios.

When undergraduate students in the intervention study (chapter 5) were presented with either the bacteria or plant scenario, majority chose line graph with 55% in Spring 2015 and 53% in Spring 2016 (Figure 5.2). This finding is consistent with that of the professors in the interview, think-aloud study (chapter 2) and graphs chosen by other undergraduate students and graduate students. Graph choice in the post survey did not change very much. The line graph was still a popular choice with 51% of the students choosing it in Spring 2015 and 55% choosing it in Spring 2016 (Figure 5.2). An interesting finding from the post surveys in both semesters was that students suggested a greater variety of graphs, as compared to the pre surveys (Figure 5.2). This suggests that repeated exposure to the guide to data displays, and to students' graphs presented in PowerPoint presentations and graph critiques could have influenced their graph choice. However, this does not mean that a box and whisker plot or a histogram are appropriate data displays for the plant or bacteria scenario (Appendix 2A). Line graphs were also a common choice across all participant groups in the think-aloud interviews (Appendix 2D). All of the expert professors (chapter 2) chose to display data from the plant or bacteria scenario in a line graph (Appendix 2D). The expert professors justified their reasoning with graph choice with the question and/or hypothesis, prior experience, and discipline-based knowledge (Figure 2.7), whereas the undergraduate students in the interview did not (Figure 2.7). As a follow up, it would be interesting to interview the students who did not choose a line graph as their graph choice for the post-survey, to understand the reasoning behind their graph choices and strategies used. A future analysis can look at how the graphs constructed by student groups over the semester influenced the graph choices and reasoning on the post survey.

Finding #4: Reflective activities with feedback in this study advance student learning about graphing

Students' written reflections on experimental concepts, research questions, and hypotheses when reflecting on graph choice need to be improved in both intervention and non-intervention semesters. Teaching students how to reflect will not only boost their self-confidence with graph construction and evaluation (Figure 5.15), but it will also emphasize the types of concepts that they should think about when reflecting and critiquing graphs in the classroom and in the media. Although we provided students with many opportunities in the non-intervention semester to reflect on their own graphs (chapter 3) -- and with having students give peer feedback in the intervention semester -- we did not give any feedback in the non-intervention semester. Instead, we gave general feedback in the intervention semester, because we were limited with time in the classroom. Although the amount of feedback students received could have factored into their improvement with graphing, we did not directly measure this variable in our study. Past research in learning psychology has shown that either immediate or delayed instructor feedback, as compared to no feedback, enhances students' learning (Butler & Roediger, 2008). Since graphing was a secondary component in the physiology laboratory, we had to prioritize what we explicitly gave feedback on, though we were always available in the lab and provided students feedback as needed. Future research questions aimed at instructor feedback on student's graph reflections and in-class feedback should be addressed.

6.3 Scope of Future Research

This dissertation provides evidence towards understanding the reasoning gaps between expert and novice biologists in clinical and classroom contexts and offers solutions to remediate these difficulties, it sets a foundation for potential future research in this ongoing, interdisciplinary field of research. Some avenues of future research are highlighted below.

6.3.1 Reasoning with Graphing in Different Contexts

The data presented in this thesis were collected from expert and novice biologists at a single Midwestern United States research university, which is a unique environment with its own curriculum and student population, which may be different from liberal arts institutions, and community colleges. Furthermore, since our study consisted of a small group of participants, the claims we presented are not broad generalizations to the types of things that all experts or novices do or think. We are stating our findings as illustrated by ten undergraduate students with no research experience (UGNR), five with research experience (UGR), eight graduate students (GS), and five professors. However, many of our findings are consistent with and extend from previous work by others in K-12 (Brasell & Rowe, 1993; Ainley, 1995; Tairab & Al Naqbi, 2004) and undergraduate (Picone et al., 2007; Leonard & Patterson, 2004; Bray-Speth et al., 2010; McFarland, 2010) contexts. To verify our findings fully, future work is needed at other types of institutions, in different disciplinary fields, and with their own unique participants in order to fully understand and appreciate what the reasoning is like for graph choice and construction.

In order to gain a finer grained appreciation on the expertise continuum and to understand when, how, and why reasoning with graph difficulties arise, one future avenue could use the methods listed in chapter 2 and expand to include post-doctoral candidates and junior faculty. Expanding the sample size to include late elementary, middle school, high school students and their teachers could potentially serve two or more purposes. First, it would allow us to track students over time to see how students reason with graph choice, construction, and will allow us to diagnose any misconceptions early on in the learning process. Second, by interviewing elementary school teachers, we can modify the evidence-based graphing materials to fit their needs. Third, by interviewing both students and their teachers will allow us to make appropriate modifications to the graphing materials for students at that cognitive level, which can be modified as the students advance through school.

Another possibility for future studies about graph choice and construction reasoning is to expand to different types of data sets and disciplines. For the purpose of our study, we provided all participants with a simple dataset from an experiment in biology with one independent variable, one dependent variable, two treatments, with three replicates each. In order to replicate our study but 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 UGR 3 showed difficulty with scaling axes (Padilla, McKenzie & Shaw, 1984; Li and Shen, 1992; Brassell & Rowe, 1993; Ainley, 2000) as indicated by the awkward positioning of the axis breaks, and UGNR3 showed difficulty with variables, as indicated by their graph (Tairab & Al-Naqbi, 2004; Appendix 2D). The simplicity of the dataset may have caused Professor 2 to go into automatic instructor mode, and may have reminded the professor of their lecture conditions where they are pressed for time in a 50 minute lecture and to quickly convey the take home message, they quickly sketch the data instead of plotting it.

Lastly, participants in this study constructed graphs manually using a LiveScribe pen and paper, instead of using the modern and conventional method of graph construction on the computer. Having participants narrate their thought process during manual construction allowed us to understand their reasoning fully. If we had asked participants to construct graphs using software programs, it may have tampered with their graph choice by biasing them towards graph choices presented by the software package. By using manual construction, we were able 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 & Bowen, 2003). We saw evidence of this with our professor population, Professor 2 in particular, who studied the data table, then sketched the data with error bars in order 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 reveal areas of difficulties and competencies with graphing further.

6.3.2 Expand and Extend Graphing Materials

These graphing materials have the potential to be part of the quantitative literacy conversation, starting at the K-12 levels (NGSS, 2012) and bridging into the undergraduate (AAMC-HHMI Committee, 2009; AAAS, 2011) and graduate biology curricula. By having students practice using these materials early on in elementary school and adjusting the graphing materials according to students' academic and cognitive development levels, we can provide students with appropriate scaffolds and guidance with graphing. This will also allow instructors to diagnose student difficulties with various aspects of graphing as they appear, so that by the time students graduate from college, they have expertise in all components of MRC, allowing them to understand the function of graphs, inventing graphs, critique graphs, and reflect on graphs. Future students should explore the general utility of the graphing materials in a variety of classroom contexts and with students at various levels in their undergraduate education. Additionally, external validation of these graphing materials in biology and non-biology classrooms and with students of different demographics and cognitive levels may reveal novel uses of that may have been overlooked in this classroom study.

The scope of this dissertation is limited to purely that of biological sciences. Since graphs are found in many different STEM and non-STEM fields, it would be interesting to see how these materials are used by students in other STEM disciplines and by people in the business administration field. This would be a significant step toward improving graphing, since past research studies suggest similar difficulties to graph choice (e.g. Leonard & Patterson, 2004, agriculture), construction (e.g. Brasell & Rowe,1993 physics), and communication (Schriger & Cooper, 2002, medicine).

6.3.3 Paper and Pen VS Computers

In chapter 2, we had the participants create graphs using the Livescribe pen and paper method because we wanted to see the reasoning behind graph choice and construction. In chapter 3 and 5, in the classroom study, students worked in groups to create graphs. Although in chapter 5 we implemented CAM and the three graphing materials: step-by-step guide, guide to data displays, and the graph rubric, and since we weren't micromanaging the student groups, we don't know how the materials were used.

6.3.5 Group Dynamics

In chapters 3 and 5 of this dissertation, we presented graphing studies in an upper-level undergraduate physiology laboratory setting. In these settings, students worked in groups of four to design experiments, collect data, and present their findings to the classroom in a PowerPoint. One major component of the PowerPoint is a graph or multiple graphs that students made of their data. Since we did not micromanage the groups, note how the graph materials were being used, or survey the students to see who in the group was in charge of graph construction, we do not have this information. Additionally, since student groups worked outside of lab time on their presentations, we do not know who made the graphs, and the thought processes that went into graph choice and construction. To understand group dynamics and usage of graph materials in the classroom, we can video tape student groups to see how they would construct a graph of the data they just collected. This would allow us to see if it is only one group member who is the dominant player in the graph or if multiple students have input to how the graph looks. It would also be interesting to see if the student who takes charge also takes the time to teach the other group members, explain the advantages, disadvantages, and usage of the graph to the group members, or if the other group members just let the lead student take charge and do not question his ideas or methods. Additionally, it would be interesting to see how students influence each other's learning. This would allow us to understand if students from the same group display similar difficulties with graph choice and construction or if their are independent from the group's ideas.

6.3.6 Automated Instructor Feedback on Graphing

In chapters 3 and 5, students either provided written reflections or critiques for their graphs. In chapter 3, the instructors did not give students feedback on their graphing reflections and in chapter 5, the instructors gave the entire lab section general feedback on their graphs, due to time limitations. Past research has shown that targeted feedback is an important component when training novices to become experts (diSessa, 2000) and feedback improves student learning (Butler & Roediger, 2008) One way to allow for quick and constructive feedback to the students, especially in a large classroom setting, is by writing a program that can automatically grade student reflections, or using the Calibrated Peer Review software. Although automatic grading software has its limitations, it is a good first step to encourage more instructors to enforce their students to engage in reflective exercises (increase MRC). Furthermore, the grading software has the potential to provide students with expert feedback that incorporates themes from the clinical interviews like experimental concepts and aligning the graph with the research question and hypothesis. This will not only help students improve their graphing reflections, but will also help them with critiquing graphs encountered outside of the classroom.

6.4 Impact of this Dissertation

The ability to apply higher order reasoning skills such as constructing, interpreting, and critiquing graphs is part of being scientifically literate and a vital skill for biology undergraduate students, as well as all other students in STEM and non-STEM disciplines to master (AAAS, 2011; Gormally et al., 2012). In this dissertation we report reasoning with data, graph choice and construction with professors, graduate students, and undergraduate students in a clinical interview setting. These data informed the development of the step-by-step guide, to aid novice students when they think about data, construct graphs, and reflect on the graph choices. Our second study with undergraduate students in an upper-level physiology lab informed the development of the guide to data displays, that explains the purpose, advantages, and disadvantages of each type of graphical representation. Lastly, the graph rubric was informed by pre-existing literature and our previous work on graphing. These three materials when implemented together in a classroom show improvements in students' graph choice and construction. These materials have only been tested in the context of undergraduate biology and future work can focus on testing their impact at other grade levels and in other contexts.

References

AAMC-HHMI Committee. (2009). Scientific foundations for future physicians. *Washington, DC: Association of American Medical Colleges*.

Ainley, J. (1995). Re-viewing graphing: Traditional and intuitive approaches. For the Learning of Mathematics, 15(2), 10-16.

Ainley, J. (2000). Transparency in graphs and graphing tasks: An iterative design process. *The Journal of Mathematical Behavior, 19*(3), 365-384.

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

Bakker, A. (2004). Reasoning about shape as a pattern in variability.*Statistics Education Research Journal*, *3*(2), 64-83.

Barraquand, F., Ezard, T. H., Jørgensen, P. S., Zimmerman, N., Chamberlain, S., Salguero-Gómez, R., Curran, T.J.,& Poisot, T. (2014). Lack of quantitative training among early-career ecologists: a survey of the problem and potential solutions. *PeerJ*, *2*, e285.

Bodner, G. M. (2006). The role of theoretical frameworks in chemistry/science education. ln. Theoretical frameworks for research in chemistry/science education, 2-26J.

Bodner and Orgill (2007). The role of theoretical frameworks in chemistry/science education.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). How people learn.

Brasell, H. M., & Rowe, M. B. (1993). Graphing skills among high school physics students. School Science and Mathematics, 93(2), 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*(3), 323-332.

Butler, A. C., & Roediger, H. L. (2008). Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. *Memory & Cognition*, *36*(3), 604-616.

Collins, A., Brown, J. S., & Holum, A. (1991). Cognitive apprenticeship: Making thinking visible. *American educator*, *15*(3), 6-11.

Duffy, T. M., & Jonassen, D. H. (1992). Constructivism: New implications for instructional technology. Constructivism and the technology of instruction: A conversation, 1-16.

Gormally, C., Brickman, P., & Lutz, M. (2012). Developing a Test of Scientific Literacy Skills (TOSLS): measuring undergraduates' evaluation of scientific information and arguments. *CBE-Life Sciences Education*, *11*(4), 364-377.

Leonard, J. G., & Patterson, T. F. (2004). Simple computer graphing assignment becomes a lesson in critical thinking. NACTA journal, 48(2), 17-21.

Li, K. Y., & Shen, S. M. (1992). Students' weaknesses in statistical projects. *Teaching Statistics*, *14*(1), 2-8.

Mevarech, Z. R., & Kramarsky, B. (1997). From verbal descriptions to graphic representations: Stability and change in students' alternative conceptions. *Educational Studies in Mathematics*, *32*(3), 229-263.

McFarland, J. (2010). Teaching and Assessing Graphing Using Active Learning. *MathAMATYC Educator*, *1*(2), 32-39.

Monteiro, C., & Ainley, J. (2007). Investigating the interpretation of media graphs among student teachers. In *International Electronic Journal of Mathematics Education*.

National Science Teachers Association. (2012). Next generation science standards. *Retrieved on*, *30*, 12.

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*(1), 20-26.

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.

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

Roth, W. M., & Bowen, G. M. (2003). When are graphs worth ten thousand words? An expert-expert study. *Cognition and Instruction*, *21*(4), 429-473.

Redish, E.F. (2002). Teaching with the Physics Suite. Accessed September 6th, 2016: http://www2.physics.umd.edu/~redish/Book/

Tairab, H. H., & Khalaf Al-Naqbi, A. K. (2004). How do secondary school science students interpret and construct scientific graphs?. *Journal of Biological Education*, *38*(3), 127-132.

VITA

VITA

AAKANKSHA ANGRA
Purdue University
Department of Biological Sciences
915 West State Street
West Lafayette, IN 47906

Education

2016 September, PhD, Biology Education Research
Department of Biological Sciences, Purdue University, West Lafayette, IN
Dissertation: Understanding, evaluating, and diagnosing undergraduate student difficulties with graph choice and construction.
Advisor: Dr. Stephanie M. Gardner
Committee Members: Dr. Nancy Pelaez, Dr. Edward Bartlett, Dr. Signe Kastberg

2012 May,BS, Applied Biology with Research Option (Graduated with Honors)
Department of Biology, Georgia Institute of Technology, Atlanta, GA
Honors Thesis: Escape Behavior in Temora longicornis when exposed to Karenia brevis and Alexandrium fundyense.
Advisor: Dr. Jeannette Yen

Certificates

2016 April, Graduate Instructional Development Certificate
Center for Instructional Excellence, Purdue University, West Lafayette, IN

2016 April, Graduate Teacher Certificate
Center for Instructional Excellence, Purdue University, West Lafayette, IN

2015 June, Certificate in Applied Management Principles
Krannert School of Management, Purdue University, West Lafayette, IN

Curriculum Design and Research Experience

2012 – 2016, Graduate Research Assistant
Department of Biological Sciences, Purdue University
Advisor: Dr. Stephanie M. Gardner
- Expertise in social science and education research methodologies
- Conducted think aloud interviews with expert professors, novice students in the biological sciences to understand their reasoning with data, graph choice, and construction.
- Designed surveys using Qualtrics software to capture, track and analyze data
- Designed curricula materials to remediate student difficulties with graph choice and construction: graph rubric, step –by- step guide to data communication, and guide to data displays
-Tested these materials in an upper division biology laboratory classroom
-Used Microsoft Excel to organize and track progress of learning materials
-Used SPSS and SAS for rigorous quantitative analysis

2014-2015, HHMI Course Module Developer
Department of Biological Sciences, Purdue University
Advisor: Dr. Edward Bartlett
-Collaborated with various professors and assisted with development, writing, and preparation of course modules targeted at students in higher education

2013, Graduate Research Assistant, Lab Rotation
Department of Biological Sciences, Purdue University
Advisor: Dr. Edward Bartlett
- Analyzed microscopic images to quantify the density of neurons in the dorsal, ventral, and medial regions of the medial geniculate body in young and aged rats

2011, Undergraduate Biology Lab Curriculum Development Assistant
School of Biology, Georgia Institute of Technology
Advisor: Dr. Cara Gormally
- Designed and tested an inquiry-based biology lab module for the undergraduate organismal biology students titled: Sensory Ecology: Photoreception in Artemia salina
- Trained other undergraduate and graduate teaching assistants in the proper use of this lab
- As of fall 2016, the lab is still being used by students enrolled in the organismal biology course

2011 – 2012, NASA Education Student Assistant and Program Coordinator
School of Biology and Biochemistry, Georgia Institute of Technology
Advisors: Dr. Loren Dean Williams and Dr. Eric Gaucher

- Worked with high school teachers and professors at Georgia Institute of Technology to design curricula materials that aligned with Georgia high school state science standards
- Tested materials and published them online
- Assisted in teaching week-long astrobiology camp to high school students

2009-2012, Undergraduate Research Assistant
School of Biology, Georgia Institute of Technology
Advisor: Dr. Jeannette Yen
- Maintained phytoplankton and zooplankton cultures for ongoing experiments
- Conducted experiments to understand the feeding, mating, and escape behavior of copepod, Temora longicornis when exposed to harmful algal blooms Karenia brevis and Alexandrium fundyense.
- Performed data analysis
- Project served as part of the Undergraduate Research Option

Teaching Experience

Purdue University
2013-2016, Spring, Graduate Teaching Assistant Principles in Physiology Lab
Professors: Dr. Stephanie Gardner and Dr. Elwood Walls
- Lead pre-lab lecture on graphing skills
- Trained students on the usage of graphing materials
- Engaged students in inquiry-based labs
- Gave students constructive feedback on their oral presentations and graphs

2012 Fall, Graduate Teaching Assistant, Anatomy and Physiology Lab
Professor: Stuart Michael
- Lead interactive weekly pre-lab lectures on relevant topics in anatomy and physiology
- Instructed students on lab safety and animal care

Georgia Institute of Technology
2011 Spring, Undergraduate Teaching Assistant Biological Principles Lab
Professor: Dr. Cara Gormally
-Lead interactive weekly lectures on relevant lab topics and lab safety
-Engaged students in inquiry-based labs
-Helped foster skill development with reading and writing scientific literature
-Gave students constructive feedback on lab reports

2010 Fall, Undergraduate Teaching Assistant Introduction to Organismal Biology
Professor: Dr. Cara Gormally
-Lead interactive weekly lectures on relevant lab topics and lab safety
-Engaged students in inquiry-based labs

-Helped foster skill development with reading and writing scientific literature
-Gave students constructive feedback on lab reports

Mentoring Experience

Purdue University
2016 Summer, Trained Undergraduate Student, **Kenya Lee** from University of North Carolina, Chapel Hill in Qualitative Research Methods and Data Analysis, Understanding Graph Critique with Undergraduate Students

2015 Spring, Trained Rotation Graduate Student, **Mozhu Li** in Qualitative Research Methods and Data Analysis, Understanding Graph Knowledge with Expert Professors and Novice Undergraduate Students

2014 Fall, Trained Rotation Graduate Student, **Brian Demong** in Graph Rubric Design and Assessment

2014 Summer, Trained Undergraduate Student, **Mark Choi** in Qualitative Interview Transcriptions

2014 Summer, Trained **Professor Micheal Kelrick** from Truman State University in Qualitative Research Methods and Data Analysis

2014 Spring, Trained Rotation Graduate Student, **Jeff Radloff** in Qualitative Research Methods and Data Analysis

Georgia Institute of Technology
2010-2012, Trained Graduate Student, **Larisa Pender-Healy** and Undergraduates, **Brooke Beaulieu** and **Erin Sentell**, on proper plankton culture and care techniques

Leadership and Service in the Community

West Lafayette, IN
April 2016, Organizer for the Spring Fest Outreach
-Designed an activity called "Bury your Trash" to teach young children and their parents about biodegradable products.
- Recruited and trained other graduate students to help run the event

2015-2016, Graduate Admissions Committee
-Work on a team with biology faculty to review and evaluate applications for incoming graduate students, conduct interviews, and award scholarships

2015-2016, President for Discipline Based Education Research for Graduate Students
- Serve as the primary contact between the group and the University
- Submitted G-SOGA grants (Fall 2015 and Spring 2016) to acquire funding
- Invite speakers
- Organize and facilitate bi-monthly meetings
- Appoint officers and delegate tasks

2015-2016, News Correspondent for the Purdue International Biology Education Research Group (PIBERG)
-Update the PIBERG website with current events

2015 April, Lead organizer for the Biology Graduate Student Retreat
-Worked in a team to organize a full day graduate retreat
-Responsibilities consisted of obtaining funding, advertising, scheduling the venue, caterer, and facilitating a relaxing environment for the retreat

2015 April, Graduate Student Panel
-Served on a career panel for undergraduate biology students at Purdue University

2014-2015, Secretary for the Biology Graduate Student Council
-Responsibilities consisted of taking notes at the meeting
-Serving as a main source of communication between the executive board of members, graduate students, post-docs, and faculty

2014 Summer, Graduate Student Panel

2014 Spring, Co-organizer for the Prospective Graduate Interview Weekend
-Worked in a team to organize events for incoming graduate students

2013 Fall, Purdue International Biology Education Research Group Co-coordinator
-Worked with Dr. Stephanie M. Gardner to organize PIBERG Meetings, speakers, lead discussions

2013-2015, Big-sister/Big-brother biology program
-Assisting first semester graduate students in the biology department with questions relating to academic and campus culture, and serving as a role model

2013-2015, Lead-organizer of Graduate Research in Progress Series
-Worked in a team to organize monthly research presentations
-Responsibilities consisted of advertising for events, establishing email correspondence with the speaker, and designing a customized feedback rubric for the speaker

2013-2016 Spring, Blue Ribbon Panel Judge for the Lafayette Regional Science and Engineering Fair
-Judged various discipline specific projects by elementary and middle school students
-Worked on a team of judges to recognize and award outstanding projects

2013 Spring, Judge for the Mini Meds Weekend Program
-Judged projects on the human body systems for elementary school children
-Worked on a team of judges to recognize and award outstanding projects

Atlanta, GA
2004- 2012, Instructor and Co-founder of the Science Night Out Program at Fernbank Science Center
-Designed and implemented exciting hands on activities for elementary school children on a monthly basis
-Lead all classroom activities and explained the science behind each activity in a simplified manner

Grants, Awards, and Honors

2016, Bilsland Dissertation Fellowship, Purdue University
-Competitive award given to senior graduate students to complete writing their dissertation
-Awarded $11,000, but declined

2016, 1st Place Sigma Xi Poster Award, Purdue University
Title: Using the Cognitive Apprenticeship Model to Improve Undergraduate Biology Students' Graph Choice and Construction
-Awarded $200

2016, Graduate Student Organization Grant Allocation, Purdue University
-Competitive award to help fund professional development and social activities for the Discipline Based Education Research Group for Graduate Students
-Awarded $750

2015, Andrews-John Scott A. and the Ella A. Travel Award, Purdue University
-Competitive departmental travel award to help students with conference expenses
-Awarded $950

2015, 2nd Place Oral Presentation Award, Purdue University
Title: Development, Validation, and Instructional Implications of the Graph Rubric
-Awarded $100 to use towards research expenses

2014, 2nd Place Poster Award, Purdue University
Title: Using Student Laboratory Reflections to Reveal Reasoning behind Graph Choice and Construction
-Awarded $40 to use towards research expenses

2014, Women in Science Program Travel Award, Purdue University
-Competitive university-wide travel award to help students with conference expenses
-Awarded $500

2013, Graduate Student Society for the Advancement of Biology Education Research (SABER) Travel Award, University of Minnesota
-Competitive travel award offered to graduate students to attend the SABER conference
-Awarded $500

2012, Graduate Student Society for the Advancement of Biology Education Research (SABER) Travel Award, University of Minnesota
-Competitive travel award offered to graduate students to attend the SABER conference
-Awarded $500

2011, Kudravi Undergraduate Teaching Assistant Award, Georgia Institute of Technology
- Nomination by a faculty member or student to an undergraduate teaching assistant who demonstrated innovative teaching techniques for furthering student learning in the School of Biology
-Awarded $300

2010, President's Undergraduate Research Award, Georgia Institute of Technology
-$1000 awarded to fund student salary for summer research

2011, Faculty Honors, Georgia Institute of Technology
-Honor awarded to students for achieving a 4.0 GPA

2008-2012, Dean's List, Georgia Institute of Technology
- Honor awarded to students for achieving at least a 3.0 GPA.

2008-2012, Hope Scholarship
-Scholarship awarded to students with U.S. citizenship and permanent residents of the state of Georgia who are enrolled in an accredited state undergraduate institution
-Students must maintain an average college GPA of 3.0 or higher

Refereed Journal Articles

-Lasley-Rasher, R. , Nagel, K., **Angra, A**. and Yen, J. (April, 2016). Intoxicated copepods: ingesting toxic phytoplankton leads to risky behavior. Proceedings of the Royal Society B.
Abstract: Understanding interactions between harmful algal bloom (HAB) species and their grazers is essential for determining mechanisms of bloom proliferation and termination. We exposed the common calanoid copepod, Temora longicornis to the HAB species Alexandrium fundyense and examined effects on copepod survival, ingestion, egg production and swimming behaviour. A. fundyense was readily ingested by T. longicornis and significantly altered copepod swimming behaviour without affecting copepod survival or fitness. A. fundyense caused T. longicornis to increase their swimming speed, and the straightness of their path long after the copepods had been removed from the A. fundyense treatment. Models suggest that these changes could lead to a 25–56% increase in encounter frequency between copepods and their predators. This work highlights the need to determine how ingesting HAB species alters grazer behaviour as this can have significant impacts on the fate of HAB toxins in marine systems.

-**Angra, A**. & Gardner, S.M. (March, 2016). The Development of a Framework for Graph Choice and Construction. Advances in Physiology Education.
Abstract: Graphs are visual representations of data and play a vital role in science communication, yet past studies have shown that students in K-16 struggle with graph choice and construction. To advance graphing skills, we developed two learning and instructional materials: Guide to data displays and step-by-step guide to data communication. Here, we briefly highlight the purpose, development, and usage of these materials, which increase students' knowledge with common graphs and provide a guiding framework for data presentation.

Manuscripts in Submission and in Preparation

-**Angra, A.** & Gardner, S.M. (under review). Illustrating the expert-novice continuum in graph construction in biological sciences. Journal of Research in Science Teaching

-**Angra, A.** & Gardner, S.M. (in preparation). Assessing Graphical Competency in an Upper-Level Physiology Laboratory Course. CBE Life Sciences.

-**Angra, A**. & Gardner, S.M. (in preparation). Graphing Rubric to Assess Student Graphs Produced in an Upper-Level Physiology Laboratory Course. Advances in Physiology Education.

-**Angra, A.** & Gardner, S.M. (in preparation). Learning by osmosis: Using biology as a medium for teaching data collection and analysis. CourseSource.

Published Educational Materials

-**Angra A.** & Gardner, S.M. (2016). Step-by-step guide. Published in Advances in Physiology Education.

-**Angra A**. & Gardner, S.M. (2016). Guide to data displays. Published in Advances in Physiology Education.

-**Angra, A.** & Gormally, C. (2011). Sensory Ecology: Photoreception in Artemia salina. Published in the Georgia Tech Biology1520 lab manual.

-**May 2011, Undergraduate Research Thesis: Escape Behavior in Temora longicornis when exposed to Karenia brevis and Alexandrium fundyense.** http://smartech.gatech.edu/handle/1853/38817

-August 2011: **Training Package: Astrobiology Life on the Edge** for high school teachers and students.

-**Angra, A.** (2011). How to keep a lab notebook. Georgia Tech NASA High School Outreach. [http://gtastrobiology.weebly.com/uploads/9/5/1/9/9519638/notebook_teacher.pdf]

-**Angra, A**. and Snell, T. (2011). Eukaryotic Extremephiles: Rotifers. Georgia Tech NASA High School Outreach. [http://gtastrobiology.weebly.com/uploads/9/5/1/9/9519638/rotifers_teacher.pdf]

-**Angra, A**. and Greenwood, J. (2011). DNA Extraction from Extremophiles. Georgia Tech NASA High School Outreach. [http://gtastrobiology.weebly.com/uploads/9/5/1/9/9519638/dna_extraction_from_extremophiles_teacher.pdf]

-Greenwood, J. and **Angra, A**. (2011). Secret of Life: Micropipetting. Georgia Tech NASA High School Outreach. [http://gtastrobiology.weebly.com/uploads/9/5/1/9/9519638/micropipetting_teacher.pdf

-Greenwood, J. , **Angra, A**. and Wartell, R. (2011). Gel Electrophoresis. Georgia Tech NASA High School Outreach. [http://gtastrobiology.weebly.com/uploads/9/5/1/9/9519638/gel_electrophoresis_teacher.pdf]

-Boyd, D., **Angra, A**. and Macaluso, P. (2011). Psychrophiles-What's Growing in Your Fridge. Georgia Tech NASA High School Outreach. [http://gtastrobiology.weebly.com/uploads/9/5/1/9/9519638/psychrophile_teacher.pdf]

-Boyd, D., Macaluso, P. and **Angra, A**. (2011). Titan's Temperature. Georgia Tech NASA High School Outreach. [http://gtastrobiology.weebly.com/uploads/9/5/1/9/9519638/titan_teacher.pdf]

Oral Presentations

-**Angra, A.** (2016, July). Revealing Meta-Representational Competence with Graph Choice and Construction in Expert and Novice Biologists. Society for the Advancement of Biology Education Research, Minneapolis, MN.

-**Angra, A.** (2015, April). Development, Validation, and Instructional Implications of the Graph Rubric. Biology Graduate Student Retreat. Purdue University, West Lafayette, IN.

-**Angra, A.** (2015, February). Assessing Graphical Competency in an Upper-Level Physiology Laboratory Course. Graduate Research in Progress Series. Purdue University, West Lafayette, IN.

-Gardner, S.M. and **Angra, A.** (2013, November). Using Student's Reflections on Graphing Choices to Improve Student Learning of Quantitative Literacy at an Upper-Division Undergraduate Physiology Course. Oral presentation. 2013 Departmental Swan Lake Retreat, Plymouth, IN.

-Gardner, S.M. and **Angra, A.** (2013, October). Using Student's Reflections on Graphing Choices to Improve Student Learning of Quantitative Literacy at an Upper-Division Undergraduate Physiology Course. Oral presentation. 2013 Assessment Institute, Indianapolis, IN.

-**Angra, A.** (2013, April). Anthropogenic obstacles and raptors: Is maneuverability associated with powerline collisions? Oral presentation. Ecology Lunch Seminar Series, Purdue University, West Lafayette, IN.

-**Angra, A.** (2013, March). Raptor Maneuverability in the Modern Landscape. Oral presentation. Graduate Research In Progress Series, Purdue University, West Lafayette, IN.

Poster Presentations

-**Angra, A.** and S.M. Gardner. (2016, March).Using the Cognitive Apprenticeship Model to Improve Undergraduate Biology Students' Graph Choice and Construction. Sigma Xi, Purdue University, West Lafayette, IN.

-**Angra, A.** and S.M. Gardner. (2015, July).Development of an Analytic Rubric to Evaluate Undergraduate Student Graphs and Diagnose Difficulties in Graph Choice and Construction. Society for the Advancement of Biology Education Research (SABER). Minneapolis, MN.

-**Angra, A.** and S.M. Gardner. (2015, July).Using the Cognitive Apprenticeship Model to Improve Undergraduate Biology Students' Graph Choice and Construction. Society for the Advancement of Biology Education Research (SABER). Minneapolis, MN.

-**Angra, A.** and S.M. Gardner. (2014, July).Illustrating the expert-novice continuum in graph construction in biological sciences. Society for the Advancement of Biology Education Research (SABER). Minneapolis, MN.

-**Angra, A.** and S.M. Gardner. (2014, April). Using Student Laboratory Reflections to Reveal Reasoning behind Graph Choice and Construction. Poster presentation. Biology Graduate Student Retreat. Beck Agriculture Center, West Lafayette, IN.

-**Angra, A.** and S.M. Gardner. (2014, April). Elucidating the reasoning used by novices and experts when graphing biological data. Poster presentation. Office of Interdisciplinary Graduate Programs (OIGP) Purdue University, West Lafayette, IN.

-**Angra, A.** and S.M. Gardner. (2014, March). Elucidating the reasoning used by novices and experts when graphing biological data. Poster presentation. Annual Graduate Student Educational Research Symposium (AGSERS). Purdue University, West Lafayette, IN.

-**Angra, A.** and S.M. Gardner. (2014, February). Using Student Laboratory Reflections to Reveal Reasoning behind Graph Choice and Construction. Poster presentation. Sigma Xi Poster Competition. Purdue University, West Lafayette, IN.

-**Angra, A.** and S.M. Gardner. (2013, July). Elucidating the reasoning used by novices and experts when graphing biological data. Poster presentation. Society for the Advancement of Biology Education Research (SABER). Minneapolis, MN.

-**Angra, A.** Dowell, A., Greenwood, J., Boyd,D., Mason,T., Macaluso,P., Randall,R., Prickett,C., Anderson, E., Jones,B., Cola,J., Zimmerman,C., Gaucher,E., Snell,T., L.D.

Williams. (2012, July). Astrobiology Camp: The 5-E model. Poster presentation. Society for the Advancement of Biology Education Research (SABER). Minneapolis, MN.

-**Angra, A**., Dowell, A., Greenwood, J., Boyd,D., Mason,T., Macaluso,P., Randall,R., Prickett,C., Anderson, E., Jones,B., Cola,J., Zimmerman,C., Gaucher,E., Snell,T., L.D. Williams. (2012, July). Astrobiology Stamps: Connecting Creativity and Science in High School Curricula. Poster presentation. Georgia Intern Fellowships for Teachers Banquet, Atlanta, GA.

-**Angra, A**. Dowell, A., Greenwood, J., Boyd,D., Mason,T., Macaluso,P., Randall,R., Prickett,C., Anderson, E., Jones,B., Cola,J., Zimmerman,C., Gaucher,E., Snell,T., L.D. Williams. (2012, July). Astrobiology Camp: The 5-E model. Poster presentation. Georgia Intern Fellowships for Teachers Banquet, Atlanta, GA.

-**Angra, A**. Lasley-Rasher, R. and J. Yen. (2011, December). The Sweet Escape: The effects of toxic algae, Karenia brevis and Alexandrium fundyense on Temora longicornis escape behavior. Poster presentation. University Symposia, Georgia Institute of Technology, Atlanta, GA.

-**Angra, A.,** Macaluso, M Greenwood, J., Boyd,D., Randall,R., Prickett,C., Anderson, E., Jones,B., Cola,J., Gaucher,E., Snell,T., L.D. Williams . (2011, July). Eukaryotic Extremophiles: NASA Astrobiology Summer Camp 2011. Poster presentation. Georgia Intern Fellowships for Teachers Banquet, Atlanta, GA.

-Macaluso, M. **Angra, A**., Greenwood, J., Boyd,D., Randall,R., Prickett,C., Anderson, E., Jones,B., Cola,J., Gaucher,E., Snell,T., L.D. Williams. (July 2011). Astrobiology Summer Camp. Poster presentation. Georgia Intern Fellowships for Teachers Banquet, Atlanta, GA.

-**Angra, A.** Lasley-Rasher, R. and J. Yen. (2011, April). Presented at the Annual Undergraduate Research Opportunities Symposium. Poster titled: “Swim copepod swim! The effects of toxic algae, Karenia brevis and Alexandrium fundyense on Temora longicornis escape behavior”.

-**Angra, A.** Lasley-Rasher, R. and J. Yen.(2010, July). Run copepod run! An escape studying the effects of toxic algae, Karenia brevis on the Calanoid copepod, Temora longicornis. Poster presentation. Aquatic Biology Summer Research Symposium, Georgia Institute of Technology, Atlanta, GA.

-**Angra, A.** Lasley-Rasher, R. and J. Yen.(2009, July). The Effects of Harmful Algal Blooms on the Feeding and Mating behavior of the Calanoid copepod, Temora longicornis. Poster presentation. Aquatic Biology Summer Research Symposium, Georgia Institute of Technology, Atlanta, GA.

Professional Training

2014, CourseSource Workshop, University of Minneapolis
2014, ACED Bio Meeting, Purdue University
2013, The Best Practices in Undergraduate Biology Education: Midwest Regional Exchange Conference, Purdue University
2012, American Society for Biochemistry and Molecular Biology (ASBMB) Workshop, Purdue University
Ad Hoc Reviewer

CBE-Life Sciences Education
Freeman Biological Science 6th Edition
-Tips on Drawing Membranes
-Tips on Drawing Enzymes and Substrates
-Drawing J-Curves
-Drawing Sigmoidal Curves
-Drawing Food Webs
-Drawing Niche Models
-Drawing Arrows
-Drawing Nutrient Cycles

Local and Professional Membership

2012-2016, Women in Science Program
2012-2013, Visualization in Biochemistry Education Research
2012-Present, Purdue International Biology Education Research Group
2012-Present, Discipline Based Education Research Group
2012-Present, Biology Graduate Student Council
2012- Present, Society for the Advancement of Biology Education Research
2011- Present, Tri-Beta Biology Honor Society

References

Stephanie M. Gardner, PhD
Assistant Professor
Department of Biological Sciences
915 West State Street
Purdue University
West Lafayette, IN 47907
765- 496-2936
sgardne@purdue.edu

Nancy Pelaez, PhD
Associate Professor
Department of Biological Sciences
915 West State Street
Purdue University
West Lafayette, IN 47907
765-496-3261
npelaez@purdue.edu

Edward Bartlett, PhD
Associate Professor
Departments of Biological Sciences and Biomedical Engineering
915 West State Street
Purdue University
West Lafayette, IN 47907
765-496-1425
ebartle@purdue.edu