

BioVEDA Curriculum: An Approach to Link Conceptual and Quantitative Understanding of Variation During Experimental Design and Data Analysis

Jessica Dewey^{1,2†}, Jenna Hicks^{1†}, Maxwell Kramer¹, and Anita Schuchardt¹

¹Department of Biology Teaching and Learning, University of Minnesota, Minneapolis

²STEM Education Center, University of Minnesota, St. Paul

[†]These authors contributed equally to this work.

Abstract

When performing a biological investigation, experts constantly cycle between conceptual and quantitative modes of thinking, integrating their understanding in both worlds to make decisions about their work. This cycling and mixing of conceptual and quantitative thinking in biology can be difficult for students. One way to help students integrate conceptual and quantitative thinking is by focusing on a single topic that has both conceptual and quantitative aspects. We have designed a curriculum that focuses on variation in experimental design and data analysis (The Biological Variation in Experimental Design and Analysis [BioVEDA] curriculum). We chose to focus on the idea of variation because it must be understood both conceptually and quantitatively at many points throughout a biological investigation. Additionally, this ability to apply quantitative thinking to biological concepts has been prioritized in national undergraduate education policy documents. Our curriculum consists of five activities that ask students to explore the idea of variation during the design of an experiment, and the representation and analysis of data. The activities are based on worksheets and incorporate Think-Pair-Share techniques with discussions facilitated by the instructor. We have implemented this curriculum twice in an introductory biology laboratory class for undergraduate students that is taught by graduate teaching assistants (TAs). So far, we have found that both TAs and students benefit from working through this curriculum. Students significantly improved in their understanding of variation in the context of biological investigations, and TAs gained more confidence in their ability to teach this content using active-learning techniques.

Citation: Dewey J, Hicks J, Kramer M, Schuchardt A. 2020. BioVEDA curriculum: An approach to link conceptual and quantitative understanding of variation during experimental design and data analysis. *CourseSource*. <https://doi.org/10.24918/cs.2020.16>

Editor: Benjamin Martin, Stony Brook University

Received: 10/11/2019; **Accepted:** 3/19/2020; **Published:** 6/23/2020

Copyright: © 2020 Dewey, Hicks, Kramer, and Schuchardt. This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License, which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original author and source are credited. The authors affirm that they either own the copyright to or have received written permission to use the text, figures, tables, artwork, abstract, summaries, and supporting materials.

Conflict of Interest and Funding Statement: None of the authors has a financial, personal, or professional conflict of interest related to this work.

Supporting Materials: Supporting Files S1. BioVEDA – Zebrafish Heart Rate Lab Manual; S2. BioVEDA – Task A Lesson Plan; S3. BioVEDA – Task A PowerPoint; S4. BioVEDA – Task A Student Worksheet; S5. BioVEDA – Task A Data Sheets; S6. BioVEDA – Task C1 Lesson Plan; S7. BioVEDA – Task C1 Student Worksheet; S8. BioVEDA – Task C1 PowerPoint; S9. BioVEDA – Task C1 Data File; S10. BioVEDA – Task C2 Lesson Plan; S11. BioVEDA – Task C2 PowerPoint; S12. BioVEDA – Task C2 Student Worksheet; S13. BioVEDA – Task D Lesson Plan; S14. BioVEDA – Task D PowerPoint; S15. BioVEDA – Task D Student Worksheet; S16. BioVEDA – Supplementary Expressions Sheet; S17. BioVEDA – Task B Lesson Plan; S18. BioVEDA – Task B PowerPoint; S19. BioVEDA – Task B Student Worksheet; S20. BioVEDA – Task A Generic Worksheet; S21. BioVEDA – Task C1 Generic Worksheet; S22. BioVEDA – Task C2 Generic Worksheet; S23. BioVEDA – Task D Generic Worksheet; and S24. BioVEDA – Task B Generic Worksheet.

***Correspondence to:** 5-210 Moos Tower, 515 Delaware St. SE, Minneapolis, MN 55455. Email: aschucha@umn.edu.

Learning Goal(s)

- Learn to design and carry out an experiment that accounts for different sources of variation.
- Understand how summary statistics represent biological information.
- Understand how sample size impacts statistical analyses of data.
- Understand and apply the components of a t-test.
- Generate graphs that communicate results in a clear and complete manner.

Learning Objective(s)

Task A Learning Objectives:

- Recognize multiple sources of variability in biological experiments (e.g. organismal variation, measurement variation, environmental variation, etc.).
- Design ways to manage relevant sources of variation (e.g. use precise instruments, average measurements, sample a large number of individuals, use controls during experimental design, use different sampling strategies).

Task B Learning Objectives:

- Identify the salient features of graphical representations.
- Recognize that different types of graphical representations are appropriate for representing different types of data.
- Generate graphical representations of data to answer a specific experimental question.

Learning Objective(s) continued

Task C1 Learning Objectives:

- Develop mathematical expressions for summary statistics (e.g. average, standard deviation) without drawing on conventional resources (e.g. textbooks, internet, course notes).
- Apply summary statistics when representing experimental data and the variation present within that data.

Task C2 Learning Objectives:

- Derive the relationship between sample size and summary statistics by analyzing a data set.

Task D Learning Objectives:

- Predict statistical significance using graphical representations of two samples.
- Explain how summary statistics are used in the equation for a t-statistic.
- Evaluate statistical significance using a t-statistic.
- Explain the relationship between t-statistics, p-values, and conclusions about significance.

INTRODUCTION

Biological investigations have multiple phases: one first identifies a question about a biological phenomenon, then plans an experiment to address the question, collects data per the experimental plan, analyzes the data, and finally uses these analyses to draw conclusions about the biological phenomenon being studied (1). When experts move through these phases, they are constantly cycling between conceptual and quantitative worlds, even if they do not realize it. Questions such as “How many replicates should I collect given the magnitude of the effect I expect to see?” require linking of quantitative thinking (consideration of sample size) and conceptual thinking about the biological phenomenon (predicting the effect of certain experimental conditions on the measured outcome). This ability to apply quantitative skills to think about biological problems has been nationally prioritized as one of the core competencies for practice in biology (2,3). However, this integration of quantitative thinking skills with biological content can be challenging for students. Focusing on a single topic that can be explicitly linked to both biological and quantitative content can foster more coherent integration of quantitative thinking and biological concepts, which can improve students’ understanding of concepts and ability to transfer quantitative skills (4).

Variation is inherent to all biological investigations. Within every investigation, there is natural phenotypic variation in the organism being studied that is a product of genetic and environmental variation. Additionally, the investigator inevitably introduces procedural sources of variation during data collection, like measurement variability due to human inaccuracies, or technical limitations of the tools used. Learning to recognize, anticipate, and account for these different sources of variation during the design phase of an investigation is an essential component of the science process skills students should develop during biology laboratory courses. Variation is also central to the statistical analyses that biologists use to make sense of their data. Descriptive statistics, like standard deviation and variance, quantify variation within a data set, and statistical tests account for this variation to allow investigators to draw conclusions. Despite variation being an intrinsic component of biological investigations, variation is seldom the target of explicit instruction

in classroom settings (5,6). Here, we describe a series of five activities that facilitate in-depth exploration of variation in the experimental design and data analysis phases of a biological investigation. This exploration allows students to develop a conceptual and quantitative understanding of variation.

Due to its wide-ranging application and importance, variation is a popular topic in curricula implemented at various levels of schooling. For example, multiple studies describe curricula implemented in primary school settings that focus on developing students’ understanding of variation in the context of statistics (7,8). These curricula allow students to explore measurement variation, graphical representations of variation, and statistical quantities that represent variation. Curricula that target undergraduate students’ understanding of variation have been developed for both classroom and laboratory settings. Thompson et al. (2011) implemented a lesson that aimed to improve students’ understanding of natural variation through the use of statistics, and how that variation is dealt with when classifying species (9). Batzli et al. (2014) focus on teaching students about phenotypic variation and how it may be impacted by both genetic and environmental variation (10). Lastly, Speth et al. (2014) describe their curriculum that aims to teach students about the origin of variation and how that acts as raw material for evolution by natural selection (11). While all of these curricula target the concept of variation, they differ in focus, and often do not ask students to think both conceptually and quantitatively about variation. Our curriculum is unique in that it asks students to explore variation conceptually and quantitatively throughout the investigative process.

Our curriculum (the Biological Variation in Experimental Design and Analysis [BioVEDA]) is divided into five activities that address variation in different components of the experimentation process: 1) accounting for different sources of variation when designing and carrying out an experiment, 2) generating mathematical expressions of summary statistics from data (mean and standard deviation), 3) examining the relationship between sample size and data spread (standard deviation), 4) exploring the relationship between sample distributions, summary statistics, statistical tests (specifically t-tests), and p-values, and 5) generating graphical representations of data that include measures of variation (Figure 1).

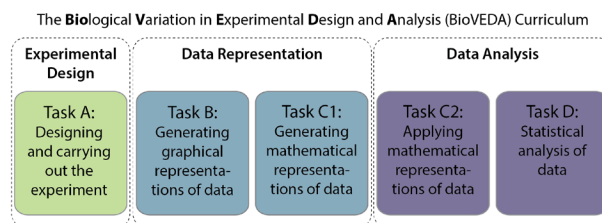


Figure 1. The BioVEDA curriculum is comprised of five activities that focus on three components of the investigative process: experimental design, data representation, and data analysis.

Intended Audience

These activities are intended for undergraduate students in introductory biology laboratory courses. We implemented these activities at a large research institution in a course that serves students majoring in biology; however, they could also be implemented at small colleges or community colleges. Approximately 200 - 400 students enroll in our introductory

biology laboratory course each semester. Students are separated into laboratory sections of approximately 20 students that meet twice a week for three hours for 15 weeks. Each lab section is taught by a graduate teaching assistant (TA), and students generally work in pairs or groups of four.

Required Learning Time

The five activities described here were implemented over a 4-week period during the first third of the semester, supplementing the regular laboratory exercises that students complete as part of the regular introductory biology laboratory curriculum. The activities range from 25 - 40 minutes in length. We present the activities in a suggested temporal order, but the activities are designed to be modular and can be customized for different experimental, course, or institutional contexts.

Prerequisite Student Knowledge

The activities were designed to be used in an introductory biology laboratory course. Students enrolled in the course have variable prior instruction in biology and statistics at our institution and we have not found a lack of biological or statistical background knowledge to negatively affect student learning or participation in these activities. However, students may benefit from basic knowledge of phenotypic variation between organisms of the same species, and from prior exposure to basic summary statistics (e.g., mean and standard deviation). Instructors who wish to use additional statistics resources with their students may find the American Statistical Association website useful (<https://www.amstat.org/>). Additionally, students will use Excel in one of the activities (Task C1). If students are not familiar with Excel, instructors may want to budget extra time to orient their students to Excel.

Prerequisite Teacher Knowledge

The instructor should have a general understanding of the types of variation that might impact a biological investigation (e.g., genetic variation, environmental variation, measurement error), an understanding of what summary statistics represent, and knowledge of t-tests. We suggest selected articles from a Nature series of reviews about statistics, p-values, and t-tests as helpful resources on these topics (12-14).

SCIENTIFIC TEACHING THEMES

Active Learning

- Think-pair-share (15).
- In-class worksheets as a platform for student discussion (16).

Assessment

The curriculum activities use two methods of assessment to evaluate student understanding. We used a multiple-choice assessment (the Biological Variation in Experimental Design and Analysis (BioVEDA) assessment, Hicks et al., in review) to measure student understanding pre- and post-implementation of the activities. This assessment asks students to apply both their conceptual and quantitative understanding of variation to answer questions about experimental design and data analysis. Readers can contact the corresponding author for a copy of the BioVEDA assessment questions. Additionally, the worksheets and corresponding whole-section discussions serve as formative assessments that allow students to check their understanding in real time.

Inclusive Teaching

A primary goal for this set of activities is to connect conceptual and quantitative thinking by exploring variation. This approach inherently includes multiple strategies to learning about the topic, a recommended approach for creating an inclusive learning environment for diverse learners (15,17). We embraced this quality by using different formats for student activities. For example, the activities begin with a whole class discussion, followed by a writing activity done in pairs, generating a pictorial representation, and generation of mathematical expressions on whiteboards in small groups. By incorporating multiple types of representations (pictorial, mathematical, graphical) and formats for students to explain their ideas (individual writing, peer discussion, whole class discussion) into the activities, we hope to broaden engagement to students who may value multiple representations or who may respond to different activity formats.

An added benefit of the formats we selected is that they are designed to foster a sense of collaboration among students. In the activities, students have discussions to select experimental design strategies used by the entire class section, pool data to explore sample size's effect on variation, and create pictorial and quantitative expression in groups, in class, with instructor feedback and discussion. All of these activities are designed with collaboration and consensus building in mind. Strategies to increase cooperation and decrease competition, like the ones used in these activities, tend to increase student achievement and feelings of engagement (18,19). This is especially important in a lab setting, where traditional lab formats may create a sense of competition among students trying to complete a prescribed experiment in parallel to other students. In addition, recent evidence has shown that collaborative lab activities help eliminate achievement gaps based on previous science experience and science career interest, which can vary widely among introductory undergraduate students for whom these activities are designed (20).

LESSON PLAN

Overview

Our curriculum consists of 5 short activities that we implemented in an introductory biology laboratory course. The activities are 25-40 minutes each and are designed to complement bench science activities that students conduct as part of the course. However, the activities described in this article do not require laboratory space or supplies, and we envision that the activities could be adapted for use in classroom-based courses in addition to laboratory-based courses. These five activities are lettered to reflect the recommended conceptual order in which we believe the activities would best be taught (i.e., Tasks A, B, C1, C2, and D; Figure 1). However, the activities are described below following the order in which we implemented them in the laboratory course to best match the established curriculum of the course. The activities are purposefully modular in nature, and though they are written here in the temporal sequence we used, the order and totality of implementation could be changed to best suit individual courses and instructor goals.

The purpose of the activities is to allow students to explore the concept of variation in the design of a biological experiment and the collection and analysis of data. We situated

the activities in the context of the laboratory experiments that students conducted as part of the normal course curriculum. Three of these activities are linked to one experiment (the zebrafish caffeine experiment, described below), one activity is situated in an experiment on primary production, and one activity is situated in an experiment on *Pseudomonas* adaptive radiation. These activities were chosen because the science process skills associated with the laboratory experiments (e.g., conducting experiments, graphing, analyzing data) complemented the BioVEDA activities. This linking of a less-familiar, abstract concept (variation) to a more familiar context (the experiment students are conducting) is intended to make the concept of variation more accessible for students. We envision that these activities can be adapted to accompany other laboratory exercises within other courses, and we elaborate on this in the Teaching Discussion.

Each laboratory session is 3 hours long and is led by a graduate teaching assistant (TA). The 25–40 minute activities can occur at any time during the laboratory session, with the exception of Task A (detailed below). Students typically work in pairs to conduct laboratory activities, so the same student pairs were used during the activities described here.

Detailed lesson plans, worksheets, and PowerPoint presentations are included as Supporting Files for each activity. The lesson plan is a step-by-step guide to implementing the activity that includes annotated pictures of student worksheets, potential student responses or ideas for certain topics, and teaching tips to help facilitate the activity. Specific PowerPoint slide numbers are referenced in the lesson plan.

Task A: Designing and Carrying out an Experiment

In this activity, we ask students to consider multiple aspects of experimental design, including controls, measurement tools, and sampling strategies. This activity is situated in a laboratory session where students perform an experiment to determine the effect of caffeine on the heart rate of zebrafish embryos. Prior to the laboratory session, we provide students with a lab manual that introduces them to the experiment and provides explicit instruction on how to perform the experiment (Supporting File S1. BioVEDA – Zebrafish Heart Rate Lab Manual). This activity is broken into two parts, 1) discussion of experimental design features, and 2) carrying out the experiment. Given that this course uses vertebrate organisms, there is a course-level Institutional Animal Care and Use Committee (IACUC) protocol that has been approved. However, students at our institution did not need special training to complete this exercise because the zebrafish embryos used in this experiment are younger than three days old. If you do not have access to a fish facility at your institution, this activity could be adapted and used with another lab where students carry out an experiment (for example, using an invertebrate like *C. elegans*). We elaborate on ways to adapt these activities in the Teaching Discussion section below.

To start this activity, ask students to share their hypotheses and predictions for this experiment in a whole class discussion (Supporting File S2. BioVEDA – Task A Lesson Plan and Supporting File S3. BioVEDA – Task A PowerPoint, slide 2). We specifically ask students to consider what a control group is, why it is necessary for the experiment, and what the control group would be for this experiment specifically. Ask 3–4 students to

share their ideas. Supporting File S2 (BioVEDA – Task A Lesson Plan) contains helpful tips for how to guide this discussion. List students' ideas on a white board to allow students to revisit these ideas after their data has been collected. This initial discussion helps students start to recognize different design features of experimental investigations. Next, ask students to describe the method (based on the laboratory manual) they will be using for collecting heart rate data during a second whole class discussion (Supporting File S1. BioVEDA – Zebrafish Heart Rate Lab Manual, Supporting File S2. BioVEDA – Task A Lesson Plan, and Supporting File S3. BioVEDA – Task A PowerPoint, slide 3). Once this method has been verbalized by students, ask them to share possible limitations of this method. Record students' ideas about the limitations of their method on a white board for later discussion.

Hand out the student worksheet for this activity after these initial discussions have been completed (Supporting File S4. BioVEDA – Task A Student Worksheet). Students should work in pairs to answer Question #1 on the worksheet. Question #1 asks students to outline two different plans for collecting 15 heart rate measurements for both their experimental and control conditions (Supporting File S3. BioVEDA – Task A PowerPoint, slide 4, Supporting File S4. BioVEDA – Task A Student Worksheet). This guides students to consider different strategies for coping with the limitations of their measurement method that they identified in the previous discussion. Once students have been given time to outline these two plans, bring the class back together and ask 2–3 groups to share and justify the strategies they developed. As groups are sharing, guide the class to a consensus on two measurement strategies for the class to use to collect their data (Supporting File S2. BioVEDA – Task A Lesson Plan and Supporting File S3. BioVEDA – Task A PowerPoint, slide 5). Supporting File S2 (BioVEDA – Task A Lesson Plan) contains tips for how to guide students to this consensus, as well as how to connect each strategy to a specific limitation of the method. Measurement Strategy A has students measure 15 embryos once each, while Measurement Strategy B has students measure five embryos three times each. To explore the potential effect of measurement strategy on their data and the trade-off between sample size and measurement accuracy, assign pairs of students to one of the two strategies described above. Once groups have been assigned, hand out the relevant data sheet to each group so students can record their measurements (Supporting File S5. BioVEDA – Task A Data Sheets). Students then answer Question #2 on the student worksheet (Supporting File S4. BioVEDA – Task A Student Worksheet) to develop a new outline of their data collection method incorporating their assigned measurement strategy.

Next, students carry out the experiment outlined in the lab manual using their assigned measurement strategy (Supporting File S1. BioVEDA – Zebrafish Heart Rate Lab Manual and Supporting File S4. BioVEDA – Task A Student Worksheet). Students will discuss differences in the data they collected and the potential effect of measurement strategy at the beginning of the next activity which occurs in the next laboratory session.

Task C1: Generating Mathematical Expressions

Task C1 links back to the experimental design and data collection that students conducted during Task A. In Task C1, students continue to build their mental model of variation by generating mathematical expressions to describe the variation

present in the data they collected. This task pushes students to think deeply about the biological information that summary statistics (e.g., mean, standard deviation) represent. This task is broken into three parts: 1) comparing and contrasting measurement strategies enacted during Task A, 2) generating mathematical expressions to describe data, and 3) using mathematical expressions to analyze data. This task requires large chart paper (or whiteboards) and markers to complete as written.

The activity starts with a Think-Pair-Share on the impact of measurement strategy on the data students collected in the previous class session (Supporting File S6. BioVEDA – Task C1 Lesson Plan). To facilitate a diverse discussion, direct students to pair with a partner who collected data using the opposite measurement strategy. Hand out Part 1 of the Task C1 worksheet (Supporting File S7. BioVEDA – Task C1 Student Worksheet), and ask students to work with their partner for 3-5 minutes to answer questions 1 and 2 (Supporting File S8. BioVEDA – Task C1 PowerPoint, slide 2). Question 1 prompts students to consider how their measurement strategy may have impacted the data that they collected in the previous activity. Question 2 asks students to consider the strengths and weaknesses of each measurement strategy. Once all students have written an answer on their worksheet, ask for 2-3 groups of students to share their answers from the worksheet questions, and target students toward the idea that different measurement methods have affordances and constraints, and that different measurement methods can impact the observed results (Supporting File S8. BioVEDA – Task C1 PowerPoint, slide 3).

Transition to the second part of the activity by asking students “How can data collected with these different measurement strategies be compared in a more meaningful way?” (Supporting File S8. BioVEDA – Task C1 PowerPoint, slide 4). Ask at least 3 students to share their ideas (possible student responses: graphs, tables, statistics, etc.), and if needed, guide students to the idea of using summary statistics to describe data sets. Hand out Part 2 of the Task C1 worksheet (Supporting File S7. BioVEDA – Task C1 Student Worksheet) and large chart paper or whiteboards, and direct students to work on question 3 in pairs. Question 3 asks students to develop and write down mathematical expressions for two different summary statistics they can use to describe their data. Students should write their mathematical expressions on the large chart paper using markers to facilitate sharing with their peers. At first, many students may be unsure of how to proceed with developing their own math expressions. See Supporting File S6 (BioVEDA – Task C1 Lesson Plan) for tips on how to support students as they work through this problem. After all student pairs have written their mathematical expressions on their large chart paper, ask students to present their expressions to either the whole class or to 2-3 nearby pairs of students. Facilitate a whole-class discussion by asking 2-3 pairs of students who have written different versions of mathematical expressions to explain their mathematical expression and justify its utility. After students have shared their expressions, wrap up the discussion by presenting the mathematical expressions that are conventionally used to describe data sets (Supporting File S8. BioVEDA – Task C1 PowerPoint, slide 6). Students should understand that the mean is a measure of central tendency, and the standard deviation is a measure of how far individual

data points typically fall from the mean.

The third part of the activity asks students to use the summary statistics (mean, standard deviation) to analyze data (Supporting File S6. BioVEDA – Task C1 Lesson Plan). Direct students to open the data sheet Excel file, which contains zebrafish heart rate data that has been collected using either Measurement Strategy A (15 fish, measured 1 time each) or Measurement Strategy B (5 fish, measured 3 times each) (Supporting File S8. BioVEDA – Task C1 PowerPoint, slide 7 and Supporting File S9. BioVEDA – Task C1 Data File). Students should answer questions 5 and 6 in pairs (Supporting File S7. BioVEDA – Task C1 Student Worksheet). Question 5 asks students to calculate summary statistics for the data set in the Excel file, and question 6 asks students to think about how the two measurement strategies may have impacted the data that was collected. Facilitate a whole-class discussion by asking 2-3 pairs of students to share their answers on question 6 (Supporting File S8. BioVEDA – Task C1 PowerPoint, slide 8). After multiple student pairs have shared, summarize the main points of the discussion using Supporting File S8 (BioVEDA – Task C1 PowerPoint, slide 9). Students should recognize that the mean values are similar between the different measurement strategies, and that spread values are more varied between the different measurement strategies.

Task C2: Applying Mathematical Representations of Variation

This task asks students to explore the relationship between summary statistics and sample size. The lesson plan in Supporting File S10 (BioVEDA – Task C2 Lesson Plan) details a step-by-step plan for implementing the activity. Begin the activity by linking back to the mathematical expressions for the mean and standard deviation that students developed in Task C1 (Supporting File S11. BioVEDA – Task C2 PowerPoint, slide 1). Hand out Part 1 of the Task C2 Worksheet (Supporting File S12. BioVEDA – Task C2 Student Worksheet), and direct students to work on Question 1, which asks students to describe the relationship between sample size and the standard deviation of a data set (Supporting File S11. BioVEDA – Task C2 PowerPoint, slide 2). Students first work on Question 1 individually, and then facilitate a discussion on students’ ideas by asking 2-3 students to share their responses. We ask students to clarify or further explain their ideas by drawing a graph that shows the relationship between sample size and data spread.

After students have explained their prior ideas about the relationship between sample size and standard deviation, hand out Part 2 of the Task C2 Worksheet (Supporting Files S12. BioVEDA – Task C2 Student Worksheet and Supporting File S11. BioVEDA – Task C2 PowerPoint, slide 3). Students should work on Question 2 individually, and then in pairs. Question 2 presents students with a data table showing the mean and standard deviation of zebrafish heart rates for increasing sample sizes and asks students to describe the relationship between sample size and standard deviation based on the data presented in the table. Facilitate a discussion on the relationship between sample size and summary statistics by asking 2-3 student pairs to share their answers on Question 2, and to justify their answers using the information in the table. After several students’ ideas have been heard, project the graph on Supporting File S11 (BioVEDA – Task C2 PowerPoint), slide

4 that shows a graphical depiction of the data in the table and a summary of the take home point of this activity: increasing sample size provides a more accurate estimate of the true mean and amount of variation in the population.

The last section of Task C2 introduces students to frequency distributions as a way to visualize sample and population data. This primes students to better understand the content of the upcoming task, which uses frequency distributions as a way to conceptualize statistical tests. To transition, ask “What do we mean by ‘true mean and variation?’” Then show a frequency distribution of the heart rate of a theoretical population of zebrafish, which is illustrated as a normal distribution (Supporting File S11. BioVEDA – Task C2 PowerPoint, slide 5). We ask “Is it possible to measure every zebrafish in the world?” and students intuitively understand that this is not possible, which pushes them to understand the need for sampling. Show frequency distributions of zebrafish heart rates generated from sample sizes of 10 (too small to give an accurate representation of the population) (Supporting File S11. BioVEDA – Task C2 PowerPoint, slide 6), 400 and 600 fish (large enough to give a fairly accurate representation of the population) (Supporting File S11. BioVEDA – Task C2 PowerPoint, slide 7) to illustrate how small sample sizes give poor estimates of the population. At the end of this lesson, students should understand that 1) every population has a true mean and amount of variation, 2) we use sample populations to estimate these values, and 3) sample size impacts the accuracy of estimates of the population mean and variation.

Task D: Statistical Analysis of Data

In this activity, students combine their knowledge of summary statistics and distributions to make sense of t-tests and p-values. This activity is broken into four parts: 1) connect sample distributions of data to running a statistical test, 2) discuss the t-test and its components, 3) introduction to p-values, and 4) application of knowledge about p-values and the t-test output.

To begin this activity, connect back to the previous discussion about population and sample distributions held during Task C2 (Supporting Files S13. BioVEDA – Task D Lesson Plan, Supporting File S14. BioVEDA – Task D PowerPoint, slide 1). Previously, students were only introduced to these ideas. In this activity, distributions will be used to explain why we need to run statistical tests. First, explain to students that when an experiment is run, multiple sample distributions are produced, and we want to determine if the control and treatment sample distributions are from the same or different populations (Supporting File S13. BioVEDA – Task D Lesson Plan). Hand out Part 1 of the Task D worksheet (Supporting Files S15. BioVEDA – Task D Student Worksheet and Supporting File S14. BioVEDA – Task D PowerPoint, slide 2). Question #1 on this worksheet asks students to make judgements about how confident they are that there is a difference between two sample distributions (Supporting File S13. BioVEDA – Task D Lesson Plan and Supporting Files S15. BioVEDA – Task D Student Worksheet). Students should first work on Question #1 alone, then pair up and identify any disagreements they have. Next, lead a short discussion, asking 2-3 student pairs to share their rankings and where they may have disagreed. To guide students towards thinking about statistical tests, ask “How can we resolve any disagreements?” (Supporting File

S13. BioVEDA – Task D Lesson Plan and Supporting File S14. BioVEDA – Task D PowerPoint, slide 3). Once students have identified statistical tests, specifically a t-test, as a way to resolve disagreements, introduce and explain the concept of hypothesis testing (Supporting File S13. BioVEDA – Task D Lesson Plan and Supporting File S14. BioVEDA – Task D PowerPoint, slide 4).

In the next segment, students further explore a t-test. Hand out Part 2 of the Task D worksheet, and ask students to work on question 2 in pairs (Supporting Files S15. BioVEDA – Task D Student Worksheet and Supporting File S14. BioVEDA – Task D PowerPoint, slide 5). This question asks students to examine the expression for a t-statistic and assign meaning to each of its components. This forces students to use their knowledge of summary statistics developed in earlier activities. Once students have finished answering this question, facilitate a discussion, guiding students to develop a consensus on what each component of the t-statistic equation represents (Supporting File S13. BioVEDA – Task D Lesson Plan, summary of take-home points on Supporting File S14. BioVEDA – Task D PowerPoint, slide 6). We have provided a “Supplemental Expressions Sheet” that may be useful for helping students with this question (Supporting File S16. BioVEDA – Supplementary Expressions Sheet).

In the third segment of this activity, students are introduced to the concept of p-values. This part of the activity is designed as a mini-lecture that does not go into detail about the statistics behind p-values. We implemented our activities in an introductory biology laboratory course, so the goal was not to (re)teach statistics to the students. Instead, we aimed to provide the students with the basic information they would need to make judgements about the statistical significance of data they may collect from laboratory experiments. Begin by asking the students “How do we determine if a t-statistic is statistically significant?” (Supporting File S13. BioVEDA – Task D Lesson Plan and Supporting File S14. BioVEDA – Task D PowerPoint, slide 7). This allows any students who are familiar with p-values to share their knowledge with the class. If a student identifies the p-value, you can push the student for more information by asking whether they know how to define the p-value (Supporting File S13. BioVEDA – Task D Lesson Plan). If no students are familiar with a p-value, bring up the p-value for the class and move on to its definition (Supporting File S13. BioVEDA – Task D Lesson Plan and Supporting File S14. BioVEDA – Task D PowerPoint, slide 8). While the p-value has a specific statistical definition, this is often confusing for students and hard to understand. In our PowerPoint we have provided the statistically accurate definition for a p-value, but emphasized the *functional* definition of a p-value instead as “how surprised you would be to see the difference in your data if your treatment had no effect” (Supporting File S13. BioVEDA – Task D Lesson Plan and Supporting File S14. BioVEDA – Task D PowerPoint, slide 8). This functional definition will likely be easier for students to apply when thinking about p-values. Explain the relationship between the p-value, your level of surprise, and your confidence in whether the difference you found is real (Supporting File S13. BioVEDA – Task D Lesson Plan and Supporting File S14. BioVEDA – Task D PowerPoint, slide 9). Explain the cut-off for statistical significance that is used in science (alpha level of 0.05) and how this cut off varies by field (Supporting File S13. BioVEDA – Task D Lesson Plan

and Supporting File S14. BioVEDA – Task D PowerPoint, slide 9). End the mini-lecture with an explanation of how to make conclusions about hypotheses using a p-value and the alpha level cut off, as well as the relationship between a t-statistic and a p-value (Supporting File S13. BioVEDA – Task D Lesson Plan and Supporting File S14. BioVEDA – Task D PowerPoint, slide 10).

The last segment of this activity asks students to combine all of their knowledge and apply it to two different worksheet questions (Supporting File S15. BioVEDA – Task D Student Worksheet). Hand out Part 3 of the Task D worksheet and direct students to work on questions #3 and #4 in pairs (Supporting File S13. BioVEDA – Task D Lesson Plan and Supporting File S14. BioVEDA – Task D PowerPoint, slide 11). Question #3 asks students to predict the size of the t-statistic and the p-value that would result from running a t-test on a set of data shown in a distribution graph, as well as what their statistical conclusion would be and why. Question 4 asks students to interpret the outcome of a marginally significant statistical test (p-value = 0.07). Often, students think of a p-value as a binary yes or no indicator and tend to think that if they do not get a significant p-value their experiment is inconclusive or a failure. This question targets these potential student ideas. Once students have answered these questions, choose whether to have a discussion about both questions or whether to focus on question #4. Teaching tips for how to guide students away from thinking of a p-value as binary are included in Supporting File S13 (BioVEDA – Task D Lesson Plan). A summary of the ideas students should understand at the end of this activity can be found in Supporting File S13 (BioVEDA – Task D Lesson Plan) and Supporting File S14 (BioVEDA – Task D PowerPoint, slide 12).

Task B: Generating Graphical Representations of Data

This task pushes students to identify the appropriate type of graph to use to represent a specific data set and asks students to recognize features of graphs that are helpful to clearly convey information in a graphical format. The placement of this task relative to the other tasks is flexible and is subject to instructor goals and discretion. A detailed lesson plan with step-by-step instructions for how to implement this activity is in Supporting File S17 (BioVEDA – Task B Lesson Plan). This task requires large chart paper (or whiteboards) and markers to complete as written.

Begin this activity by asking students to volunteer different types of graphs that can be used to represent data and compile a list on the whiteboard (Supporting File S18. BioVEDA – Task B PowerPoint, slide 1). Hand out Part 1 of the Task B Worksheet (Supporting File S19. BioVEDA – Task B Student Worksheet) and large chart paper or whiteboards, and direct students to work in pairs on Questions 1 and 2 (Supporting File S18. BioVEDA – Task B PowerPoint, slide 2). Question 1 shows a data table with a research objective and asks students to draw a graph that meets the objective. Question 2 asks students to justify their choice of graph type. After all students have had a chance to draw their graph on the provided chart paper or whiteboard, students should compare graphs with other students and answer Question 3 on the Task B worksheet (Supporting File S19. BioVEDA – Task B Student Worksheet), which asks students to identify features of other students' graphs that are similar or different to their graph (Supporting File S18. BioVEDA – Task B PowerPoint, slide 3). Lead a

discussion about features of graphs that clearly communicate scientific findings by asking 2-3 students to present their graph to the whole section and explain why they chose that type of graph (Supporting File S18. BioVEDA – Task B PowerPoint, slide 4). Ask students what types or features of graphs are useful in understanding the research objective and allow 2-3 students to respond. Make a list of useful features of graphs that clearly convey scientific findings on the whiteboard as ideas are generated through student discussion (example features are listed on Supporting File S18. BioVEDA – Task B PowerPoint, slide 5). Lastly, students answer Questions 4 and 5 on the Task B Worksheet, which ask for reflection on the discussion to identify best practices for constructing graphs (Supporting File S19. BioVEDA – Task B Student Worksheet).

TEACHING DISCUSSION

Overall Impact of the Lesson

Students' understanding of the topics covered in this lesson was measured by a multiple-choice assessment that we developed (Hicks et al., in review). This assessment evaluates students' conceptual and quantitative understanding of biological variation in the context of experimental design and data analysis. Students took this assessment at the beginning of their laboratory course, before taking part in any of the activities described above (pre-score), and again during the 6th week of class once all of the activities had been implemented (post-score). We implemented these activities in half of the sections within this lab course and compared students' learning gains to those of students who only participated in the regular lab exercises (which are similar to the zebrafish caffeine exercise detailed in Supporting File S1. BioVEDA – Zebrafish Heart Rate Lab Manual, designated as the "Traditional" curriculum in Figure 2). We found that students made significant gains in their understanding of these topics (Figure 2).

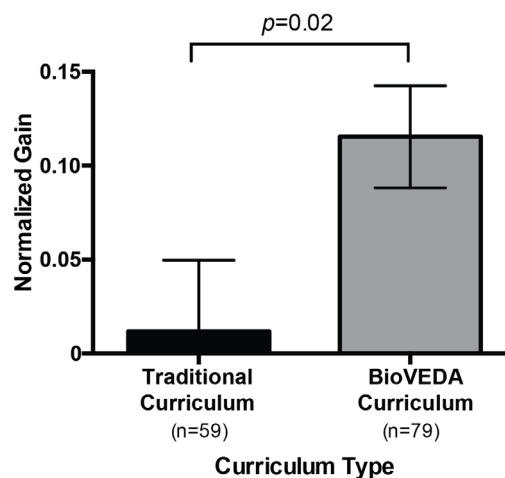


Figure 2. The BioVEDA curriculum improves students' understanding of variation in biological investigations. Normalized gain scores ((%Post-test - %Pre-test) / (100 - %Pre-test)) on the BioVEDA assessment for students who participated in the traditional course curriculum versus the traditional course curriculum with the 5 supplemental BioVEDA activities. Significance determined using a two-tailed t-test. Bar \pm error bar = Mean \pm SEM.

These activities have been implemented in two different semesters. During the first semester, JD was one of the TAs who taught in the lab course and implemented these activities. We found that students were often actively engaged with the activities and worksheets, fully participating in discussions,

and working hard with their partners in class. Occasionally students found the content challenging (e.g., when they were asked to build their own mathematical expressions), but still worked hard and participated throughout the activities. All students were able to successfully complete all of the activities.

After each semester in which we implemented these activities, we made minor changes to their organization and structure based on feedback from both students and TAs. We believe that the rendition we have described here is the most coherent and cohesive version so far.

Impacts on TAs

The activities described here were implemented in a large-enrollment laboratory course in which graduate TAs are the primary instructors. The TAs have highly variable amounts of prior teaching experience, confidence in facilitating student discussions, and content knowledge about statistics. We developed the lesson plan documents in the style of educative curricular materials used by K-12 teachers with the intent that the lesson plans would not only help TAs implement the activity but would also bolster TAs understanding of the content and pedagogy behind the lesson (21). We also conducted in-person training sessions where we modeled how the activity should be facilitated, and the TAs worked through the activities as the students. The trainings and lesson plan documents were effective in preparing TAs to facilitate the activities; all TAs were observed implementing the activities, and we believe each TA was able to faithfully reproduce the activity as we intended. We observed TAs using the lesson plans as references during class, and TAs reported using them to help prepare to teach before coming to class. We envision the lesson plans to be sufficient to prepare an instructor who is new to this content or teaching style to effectively implement the activities.

We were encouraged by TAs' responses to the content and pedagogical style of the activities, given that many of the TAs had little experience using active-learning techniques or teaching at the undergraduate level in general. We held a focus group with the TAs after each semester, and multiple TAs indicated that the interactive nature of the activities was enjoyable. One TA said, "I really appreciate this style of teaching as well, it's much more fun to lead a discussion like that rather than just telling my students a long list of things to memorize." Another TA spoke about how the activities impacted their teaching style beyond the structured activities: "I've turned all of the class discussions into a Think-Pair-Share. It seems crazy to do Think-Pair-Shares every 5 minutes, but it works!" Given that TAs successfully facilitated the activities and many responded positively to the in-person training and lesson plan documents, we believe that the curriculum could be effectively implemented in similar large-enrollment laboratory courses.

The activities were implemented such that the content covered in the activity was usually related in some way to the topic of the laboratory exercise. In rare instances, the content of the activity was fairly divorced from the content of the laboratory exercise. TAs reported that students did not respond as well to this, and that the activities ran more smoothly, and they were able to get more student buy-in when there was a clear connection between the activity and the laboratory exercise. We recommend that instructors make the links

between the activity and the classroom content explicit, if using the activities to supplement normal class instruction as we did.

Adaptations for Other Courses or Institutions

We have implemented the five activities as a whole module, however, we believe that the activities could be implemented individually, if desired. Additionally, we have implemented these activities in a specific order (described in the lesson plan), however, their design allows them to be modular and reordered as necessary. Each activity was originally designed to be topic-specific and tied to a laboratory exercise. Below, we describe ways to implement these activities in different contexts than we have described above.

Task A

This task is the only activity that directly impacts the laboratory exercise students are performing in the class. Task A asks student to consider characteristics of experimental design (e.g., measurement strategies) and how those characteristics may impact the data they collect from an experiment. This activity is situated in a lab exercise where students test the effect of caffeine on the heart rate of zebrafish embryos. The students are provided with a lab manual that outlines their method of data collection, but this activity impacts how many zebrafish embryos students are treating and observing (i.e., either 15 embryos measured once each, or 5 embryos measured 3 times each). We recognize that this may not be feasible for all courses or institutions, especially if there is no access to zebrafish embryos. However, we are confident that this activity could be implemented in any lab exercise where students could use different measurement strategies to collect data. We have generated a generic worksheet template that instructors can modify to suit their individual needs (Supporting File S20. BioVEDA – Task A Generic Worksheet). Additionally, we believe that this activity could be implemented as purely a discussion without any data collection. Discussing the potential impacts of different experimental design features is still a fruitful discussion to have with students even if they do not collect data on their own.

Task C1

This activity asks students to explore the impact of using different measurement strategies to collect data (connecting to Task A) and then generate mathematical representations of their data in order to more clearly compare the data collected using different measurement strategies. We believe that this activity could be run on its own without connecting back to the previous task. We have generated a generic worksheet template that instructors can modify to suit their individual needs (Supporting File S21. BioVEDA – Task C1 Generic Worksheet). Students could instead be provided with any kind of data and asked to compare the raw data with other students in the class. This is often difficult and can lead students to the idea of using summary statistics to compare data instead. The development of mathematical expressions is independent of context and we believe it can be implemented in conjunction with other activities or on its own. Additionally, the last part of the activity that asks students to use Excel to calculate the summary statistics for their data can be adapted for the use of a different computer program or hand calculations. We attempted to have students hand calculate their summary statistics in our first semester of implementation, however this took a lot of time and was confusing for students. If your

students are familiar with hand calculations of summary statistics, this option could work well.

Task C2

This task asks students to explore the relationship between summary statistics and sample size. This activity uses zebrafish heart rate data collected across multiple semesters of this introductory biology laboratory course. Specifically, we provide students with a table of the mean and standard deviation of heart rate data from various sample sizes (Supporting File S12. BioVEDA – Task C2 Student Worksheet) and ask them to think about how they are related. While this activity uses zebrafish heart rate data currently, it can easily be adapted for other types of data and contexts that might fit better in other courses or institutions. We have generated a generic worksheet template that instructors can modify to suit their individual needs (Supporting File S22. BioVEDA – Task C2 Generic Worksheet). The last part of the activity introduces students to the idea of population and sample distributions. We included this part of the activity in order to create a clearer transition into Task D. However, if this activity is being implemented on its own, the last section of this activity can be dropped.

Task D

This activity asks students to think about statistical tests and interpreting their output. This activity can be easily adapted for any context in which t-tests are taught, as the majority of this activity is not tied to a specific experimental context, with the exception of Questions 3 and 4 on the Task D Worksheet (Supporting File S15. BioVEDA – Task D Student Worksheet). These questions refer to an experiment to test the effect of radiation and temperature on *Pseudomonas fluorescens*. This question can be modified to align with any laboratory exercise where a t-test would be appropriate by changing the experimental context to match that of the laboratory exercise. We have generated a generic worksheet template that instructors can modify to suit their individual needs (Supporting File S23. BioVEDA – Task D Generic Worksheet).

Task B

This activity asks students to construct graphs that fulfil a specific research objective and use these graphs to identify features of graphs that help efficiently convey information. We designed this activity to accompany a laboratory exercise where students are constructing graphs of a large data set using software tools (e.g., JMP, Excel, or R). Students have to hand-draw a graph, which pushes them to make and justify decisions about information to include on their graph before moving to a computer-based graphing tool that may automate many of these considerations. We have included a research objective and a data table that draws from the data analysis that students will be subsequently performing as part of the companion laboratory exercise which focuses on productivity in grassland ecosystems (Supporting File S19. BioVEDA – Task B Student Worksheet). However, this activity could be adapted to accompany a variety of laboratory exercises by changing the research objective and provided data table. We have generated a generic worksheet template that instructors can modify to suit their individual needs (Supporting File S24. BioVEDA – Task B Generic Worksheet). This task can be successfully positioned at any point relative to the other tasks in the unit. We have implemented Task B immediately after Task C1, when students have generated their own mathematical expressions

to represent summary statistics, and also as the last activity in the sequence. We found the discussions around the activity to be equally productive in both instances and believe that instructors should arrange the tasks in the way that makes sense for their individual courses and institutions.

SUPPORTING MATERIALS

- S1. BioVEDA – Zebrafish Heart Rate Lab Manual
- S2. BioVEDA – Task A Lesson Plan
- S3. BioVEDA – Task A PowerPoint
- S4. BioVEDA – Task A Student Worksheet
- S5. BioVEDA – Task A Data Sheets
- S6. BioVEDA – Task C1 Lesson Plan
- S7. BioVEDA – Task C1 Student Worksheet
- S8. BioVEDA – Task C1 PowerPoint
- S9. BioVEDA – Task C1 Data File
- S10. BioVEDA – Task C2 Lesson Plan
- S11. BioVEDA – Task C2 PowerPoint
- S12. BioVEDA – Task C2 Student Worksheet
- S13. BioVEDA – Task D Lesson Plan
- S14. BioVEDA – Task D PowerPoint
- S15. BioVEDA – Task D Student Worksheet
- S16. BioVEDA – Supplementary Expressions Sheet
- S17. BioVEDA – Task B Lesson Plan
- S18. BioVEDA – Task B PowerPoint
- S19. BioVEDA – Task B Student Worksheet
- S20. BioVEDA – Task A Generic Worksheet
- S21. BioVEDA – Task C1 Generic Worksheet
- S22. BioVEDA – Task C2 Generic Worksheet
- S23. BioVEDA – Task D Generic Worksheet
- S24. BioVEDA – Task B Generic Worksheet

ACKNOWLEDGMENTS

We would like to thank Vanessa Pompei for her contribution in the implementation of the curriculum, and all of the teaching assistants who participated in the implementation of the new curriculum. We also thank Yaniv Brandvain for his helpful statistical advice in designing the content of the activities.

REFERENCES

1. Wild CJ, Pfannkuch M. 1999. Statistical Thinking in Empirical Enquiry. *International Statistical Review* 67:223-248.
2. American Association for the Advancement of Science. 2011. Vision and change in undergraduate biology education: A call to action. Washington, DC.
3. Council NR, Studies D on E and L, Sciences B on L, Century C on UBE to PRS for the 21st. 2003. *BIO2010: Transforming Undergraduate Education for Future Research Biologists*. National Academies Press.
4. Schuchardt AM, Schunn CD. 2016. Modeling Scientific Processes With Mathematics Equations Enhances Student Qualitative Conceptual Understanding and Quantitative Problem Solving. *Science Education* 100:290-320.
5. Batzli JM, Knight JK, Hartley LM, Maskiewicz AC, Desy EA. 2016. Crossing the Threshold: Bringing Biological Variation to the Foreground. *CBE Life Sci Educ* 15.
6. Hallgrímsson B, Hall BK. 2005. CHAPTER 1 - Variation and Variability: Central Concepts in Biology, p. 1-7. In *Variation*. Academic Press, Burlington.
7. Lehrer R, Schauble L. 2004. Modeling Natural Variation Through Distribution. *American Educational Research Journal* 41:635-679.
8. Lehrer R, Kim M, Schauble L. 2007. Supporting the Development of Conceptions of Statistics by Engaging Students in Measuring and Modeling Variability. *Int J Comput Math Learning* 12:195-216.
9. Thompson ED, Bowling BV, Whitson M, Naczi RFC. 2011. Engaging Students in Natural Variation in the Introductory Biology Laboratory via a Statistics-

- based Inquiry Approach. *AM BIOL TEACH* 73:100-104.
10. Batzli JM, Smith AR, Williams PH, McGee SA, Dósa K, Pfammatter J. 2014. Beyond Punnett Squares: Student Word Association and Explanations of Phenotypic Variation through an Integrative Quantitative Genetics Unit Investigating Anthocyanin Inheritance and Expression in *Brassica rapa* Fast Plants. *CBE Life Sci Educ* 13:410-424.
 11. Speth EB, Shaw N, Momsen J, Reinagel A, Le P, Taqieddin R, Long T. 2014. Introductory Biology Students' Conceptual Models and Explanations of the Origin of Variation. *CBE Life Sci Educ* 13:529-539.
 12. Altman N, Krzywinski M. 2015. Sources of variation: Points of significance. *Nature Methods* 12:5-6.
 13. Krzywinski M, Altman N. 2013. Points of significance: Significance, P values and t-tests. *Nature Methods* 10:1041-1042.
 14. Krzywinski M, Altman N. 2014. Points of significance: Comparing samples-part I. *Nature Methods* 11:215-216.
 15. Tanner KD. 2013. Structure Matters: Twenty-One Teaching Strategies to Promote Student Engagement and Cultivate Classroom Equity. *LSE* 12:322-331.
 16. Allen D, Tanner K. 2005. Infusing Active Learning into the Large-enrollment Biology Class: Seven Strategies, from the Simple to Complex. *Cell Biol Educ* 4:262-268.
 17. Handelsman J, Miller KS, Pfund C. 2006. *Scientific Teaching*. W.H. Freeman and Company, New York, NY.
 18. Johnson DW, Maruyama G, Johnson R, Nelson D, Skon L. 1981. Effects of cooperative, competitive, and individualistic goal structures on achievement: A meta-analysis. *Psychological Bulletin* 89:47-62.
 19. Kyndt E, Raes E, Lismont B, Timmers F, Cascallar E, Dochy F. 2013. A meta-analysis of the effects of face-to-face cooperative learning. Do recent studies falsify or verify earlier findings? *Educational Research Review* 10:133-149.
 20. Premo J, Cavagnetto A, Davis WB. 2018. Promoting Collaborative Classrooms: The Impacts of Interdependent Cooperative Learning on Undergraduate Interactions and Achievement. *LSE* 17:ar32.
 21. Davis EA, Krajcik JS. 2005. Designing Educative Curriculum Materials to Promote Teacher Learning. *Educational Researcher* 34:3-14.

Activity	Description	Time	Notes
Using mathematical expressions to analyze data.	1. Students use the Task C1 Data Sheet to use Excel to calculate summary statistics for provided data and answer Q5 and Q6 (Task C1 Worksheet, Part 2).	5 minutes	<ul style="list-style-type: none"> The Task C1 Data Sheet is provided in Supporting File S9. A detailed Lesson Plan including step-by-step instructions, PowerPoint slide number references, and anticipated student responses is in Supporting File S6. PowerPoint slides with notes are in Supporting File S8. The Task C1 Worksheet is in Supporting File S7.
	2. Instructor facilitates a discussion on whether and why measurement strategy impacted the data that was collected by asking students to share responses for Q6.	5 minutes	
Task C2: Applying Mathematical Expressions			
Preparing for Class:	Print worksheet copies.	Variable	<ul style="list-style-type: none"> Task C2 Worksheet is in Supporting File S12
Connecting back to summary statistics	Instructor gives a review of summary statistics (slide #1).	5 minutes	<ul style="list-style-type: none"> A detailed Lesson Plan including step-by-step instructions, PowerPoint slide number references, and anticipated student responses is in Supporting File S10. PowerPoint slides with notes are in Supporting File S11.
Analyzing the relationship between sample size and summary statistics	<ol style="list-style-type: none"> Students individually answer Q1 on the Task C2 worksheet. Instructor asks 2-3 students to share and justify their ideas on Q1. Students work on Q2 of the Task C2 worksheet individually, then in pairs/small groups. Instructor asks 2-3 students to share and explain their ideas on Q2. 	15 minutes	<ul style="list-style-type: none"> A detailed Lesson Plan including step-by-step instructions, PowerPoint slide number references, and anticipated student responses is in Supporting File S10. PowerPoint slides with notes are in Supporting File S11. The Task C2 worksheet is in Supporting File S12
Connecting sample size to population and sample distributions	Instructor uses PowerPoint slides #5-7 to show how sample data can be presented using frequency distributions.	5 minutes	<ul style="list-style-type: none"> PowerPoint slides with notes are in Supporting File S11.
Task D: Statistical Analysis of Data			
Preparing for Class:	Print worksheet copies.	Variable	<ul style="list-style-type: none"> Printing Parts 1, 2, and 3 of the Task D Worksheet (Supporting File S15) separately can be helpful to constrain students' focus to the relevant parts of the lesson.
Use sample distributions to practice making statistical predictions	<ol style="list-style-type: none"> Instructor connects back to discussion of population/sample distributions. Students work on Q1 of the Task D Worksheet individually, then in pairs/small groups. Instructor asks 2-3 groups to share and explain their ideas on Q1. 	8 min	<ul style="list-style-type: none"> A detailed Lesson Plan including step-by-step instructions, PowerPoint slide number references, and anticipated student responses is in Supporting File S13. PowerPoint slides with notes are in Supporting File S14. The Task D Worksheet is in Supporting File S15.

Activity	Description	Time	Notes
Identify components of a t-test	<ol style="list-style-type: none"> Instructor introduces the idea of hypothesis testing to students. Students work on Q2 of the Task D Worksheet in pairs. Instructor asks 2-3 students to share their ideas for Q2. 	10 min	<ul style="list-style-type: none"> A detailed Lesson Plan including step-by-step instructions, PowerPoint slide number references, and anticipated student responses is in Supporting File S13. PowerPoint slides with notes are in Supporting File S14. The Task D Worksheet is in Supporting File S15.
Brief Introduction to p-values	Instructor uses PowerPoint slides #7-10 to introduce students to the concept of p-values.	5 min	<ul style="list-style-type: none"> PowerPoint slides with notes are in Supporting File S14.
Application of knowledge of the t-statistic equation	<ol style="list-style-type: none"> Students work on Q3 and Q4 on the Task D worksheet in pairs. Instructor asks 2-3 pairs of students to share and explain their ideas for Q4. 	10 min	<ul style="list-style-type: none"> A detailed Lesson Plan including step-by-step instructions, PowerPoint slide number references, and anticipated student responses is in Supporting File S13. PowerPoint slides with notes are in Supporting File S14. The Task D worksheet is in Supporting File S15. If time allows, the instructor can also guide students through a discussion of Q3.
Task B: Generating Graphical Expressions			
Preparing for Class:	<ol style="list-style-type: none"> Print worksheet copies. Bring markers and large chart paper, whiteboards, or something similar so students can easily share/discuss their graphs with peers. 	Variable	<ul style="list-style-type: none"> The Task B Worksheet is in Supporting File S19.
Identify possible types of graphs	<ol style="list-style-type: none"> Instructor asks students to share types of graphs that can be used to represent data. Write list of graph types on the board. 	2 minutes	<ul style="list-style-type: none"> A detailed Lesson Plan including step-by-step instructions, PowerPoint slide number references, and anticipated student responses is in Supporting File S17. PowerPoint slides with notes are in Supporting File S18.
Construct drafts of graphs on chart paper	<ol style="list-style-type: none"> Students work on Q1 and Q2 on the Task B Worksheet. When ready, students draw their graph on the large chart paper (provided by the instructor). 	10 minutes	<ul style="list-style-type: none"> A detailed Lesson Plan including step-by-step instructions, PowerPoint slide number references, and anticipated student responses is in Supporting File S17. PowerPoint slides with notes are in Supporting File S18. The Task B Worksheet is in Supporting File S19.
Identify similarities and differences between graphs	<ol style="list-style-type: none"> Students hang up chart paper. Students work on Q3 on the Task B Worksheet. 	3-5 minutes	<ul style="list-style-type: none"> A detailed Lesson Plan including step-by-step instructions, PowerPoint slide number references, and anticipated student responses is in Supporting File S17. PowerPoint slides with notes are in Supporting File S18. The Task B Worksheet is in Supporting File S19.
Identify important features of clear graphs	<ol style="list-style-type: none"> Instructor chooses 2-3 pairs of students that created different graph types to share the type of graph they chose and why. Instructor asks students to identify features of clear graphs, and charts these ideas on the board. 	10 minutes	<ul style="list-style-type: none"> A detailed Lesson Plan including step-by-step instructions, PowerPoint slide number references, and anticipated student responses is in Supporting File S17. PowerPoint slides with notes are in Supporting File S18.