Supplemental Material

CBE—Life Sciences Education

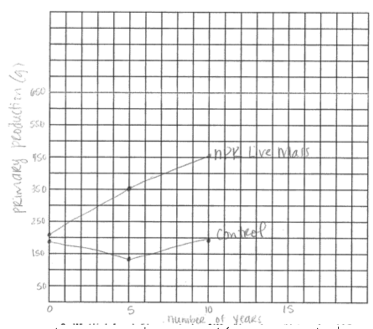
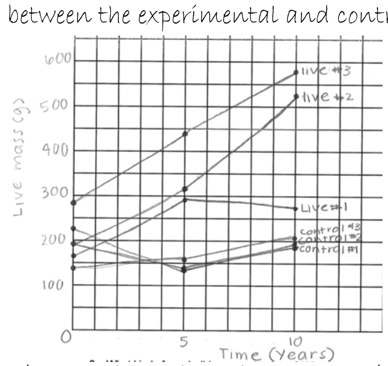
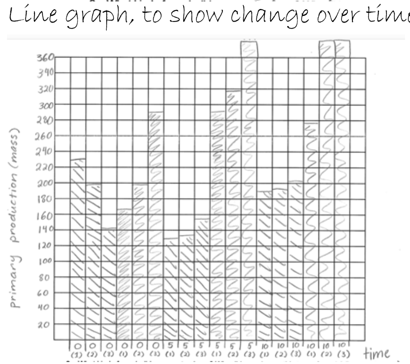
Dewey *et al*.

Table S1

Selected task instructions, example student responses and task outcomes for each module

Module	Task Instructions	Example Student Responses	Task Outcomes
A	Your goal is to decide how to collect the most reliable and accurate heart rate data possible. Outline two possible plans for collecting 15 heart rate measurements for both your experimental and comparison conditions. (Possible parameters to consider might include the number of fish to be sampled, the number of times an individual fish will be measured, etc.)	Plan A: There will be 15 control individuals & 15 experimental individuals. Each heart rate will be collected only once per individual. Plan B: There will be 5 control individuals and 5 experimental individuals. Each heart rate will be recorded 3 times per individual. We will be using the second strategy and measuring 5 individuals 3 times each. The benefit of this strategy is we can see the effect over an extended period of time and repetition helps improve accuracy in measurement.	To surface student ideas about sources of variation in the experiment. To reflect on different measurement methods with respect to accuracy and reliability. To collect data via two different measurement methods for Task C1.
C1	Refer back to the Task A worksheet data table. Write two mathematical expressions that provide summary statistics of the data you collected. Use the consensus mathematical expressions to summarize the data from the two measuring strategies. Compare your summary statistics. Do the comparison samples differ from one another? What evidence from your summary statistics supports your conclusion?	$\frac{\Sigma\ all\ data\ points}{15} = avg$ $\sqrt{\frac{(each\ data\ point - mean\ data\ points)^2}{\#\ of\ all\ data\ points}}$ Range=Maximum bpm-minimum bpm The comparison samples differ slightly from each other. In strategy B, the standard deviation is lower meaning the values are closer to the mean. Strategy B is more accurate but not by a whole lot.	To make connections between the two different measurement methods and variation in data. To develop consensus representations of mean and variation in data. To examine the effect of the two different measurement strategies on standard deviation and draw conclusions about accuracy.

C2	Examine the data table showing heart rate measurements from untreated zebrafish. The mean and standard deviation were calculated for various sample sizes. How would you describe the relationship between sample size and standard deviation? Support your answer with data from the table.	As sample size increases, the standard deviation increases. While a sample size of 3 fish has a standard deviation of 8.0, a sample size of 300 fish has a standard deviation of 31.3. The standard deviation increases with sample size, then plateaus at the largest samples.	Having made the connection between standard deviation and variation, to connect standard deviation to sample size.
D	Below are four graphs each containing two sample distributions. The distribution in gray is from samples that were treated with Compound X. Using your knowledge of distributions and summary statistics, rank these four graphs based on how confident you would be that Compound X has an effect (1=most confident). Explain your reasoning. Below is the formula to calculate a t-statistic. Describe (in a short phrase) what each boxed quantity represents.	The likelihood that two distributions would be as distinct and as far apart as in B by random chance is highly unlikely, whereas the distributions are so close and so variable in C, it would be hard to conclude anything – they share most of their data. A: Average of Sample A B: Average of Sample B C: Standard Deviation of Sample A D: Standard Deviation of Sample B Written above numerator: $Mean_A$-$Mean_B$ Written below denominator: St. Dev_A + St. Dev_B (Variance) (Variance)	To use graphical representations to support students' understanding of t-statistic and thus p-value. To relate the components of the t-statistic equation to means and standard deviation. To use the mathematical and graphical representations to support understanding that t-statistics is a comparison of the difference between the means and the extent of the variation of the samples.

D (cont.)	Below is data from an experiment. Using the two sample distributions shown on the graph, predict the size of the t-statistic and p-value that would result from a t-test by circling below. Then, predict whether the difference between the means of the two samples would be statistically significant and explain your reasoning.	T-statistic: Small; P-value: Large; Statistical conclusion: Not significant. S1: Since there is large overlap, the means won't be significantly different, thus the test will not have a large statistic and we will have a large p-value, a high chance that any significance is due to chance. S2: The difference in means being quite small led me to conclude that the p-value is large. The difference in means divided by the difference in variation seems to be a small number here.	To relate t-statistics to p-value and the probability that the difference in the means could occur by chance.
B	The data tables show the mass of plants collected from three different samples of a plot of land at Year 0, Year 5 and Year 10. One plot of land has been treated with fertilizer, the other plot has not. We are interested in exploring changes over time in response to fertilizer. Using data from the above tables, draft a graph showing how primary production changes over time. What kind of graph did you create? Why?	 We chose to draw a line graph because it showed the data in a chronological order. This graph also shows the differences between the experimental and control group.  Line graph, to show change over time. 	To graphically represent data using summary statistics and to identify key elements of graph.

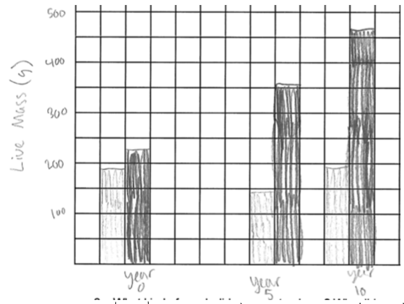
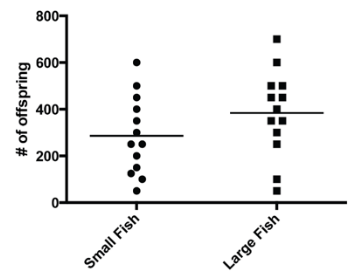
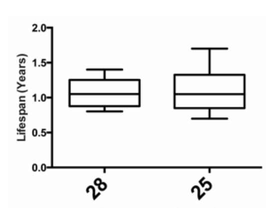
B (cont.)		I chose a bar graph so I could compare the control and experimental groups. I can look at the differences in their masses.  We created a bar graph to not only show how the NPK increased more than the control group but also to show a direct comparison between the control & NPK over each time interval.	
	After class discussion, what features should be on a graph to clearly communicate results?	Key/legend that explains color coding Title Axis labels Units Trend line Indicator of summary statistics (mean, standard deviation)	

Table S2

Sample items from the BioVEDA assessment by topic area

Investigative Phase	Topic	Sample Items
Experimental Design	Identifying sources of variation in an experiment	A researcher is interested in the effects of a drug that speeds up cell turnover on the rate of zebrafish fin regeneration. They grow two groups of 20 fish in two different fish tanks. One tank has the drug added to it, and the other does not. The researchers make identical wounds in each fish tail fin, and measure the degree of wound healing in each fish after 1 month. Which of the following is a source of organismal variation in this experiment? A. The rates of wound healing between the two treatment groups B. The size of the initial wound in the tail fin between the two groups C. **The rates of wound healing within the same treatment group** D. The size of the initial wound in the tail fin within the same treatment group
	Controlling for different sources of variation in an experiment	Jane is measuring zebrafish heart rate in response to caffeine treatment. She makes sure that the fish tanks being used for the experiment are the same size, have the same volume of water in them, and are being kept at the same temperature. Why does Jane ensure that the tank setup is similar between the caffeine and comparison groups? **A. She can be more certain that any effect she sees is due to the caffeine treatment, and not difference in housing conditions.** B. Keeping the environment constant will minimize the amount of variation between individuals in the same treatment group. C. It is good scientific practice to ensure that there is no difference in the environment between experiment and comparison groups. D. She will be able to eliminate variation between the individuals within the caffeine-treated and comparison groups.
	Understanding the relationship between sample size and genetic variation in a biological data set	A large sample size is desirable because a larger sample size will: A. Increase the amount of variation between individuals in the sample. **B. Yield a more accurate estimate of the population being sampled.** C. Decrease the amount of variation between individuals in the sample. D. None of the options are true.

Data Analysis	Representing observed variation in a data set	A group of students is testing whether large zebrafish have more offspring than small zebrafish. They collect data on the number of eggs laid by large fish and small fish. Their data is shown below (each point represents one fish, horizontal lines indicate the mean for each group).  Student A thinks that there will not be any significant difference between the number of offspring for small and large fish. Student B thinks there will be a significant difference. Who do you agree with? A. Student A, because the means are close to each other. **B. Student A, because there is a lot of variation in both samples.** C. Student B, because the means are different from each other. D. Student B, because there is less variation in small fish.
	Understanding how observed variation impacts the outcome of statistical tests	A researcher is measuring the difference in number of stripes of zebrafish. They analyze 1000 male and 1000 female fish, and find that male fish have 5 stripes and female fish have 6 stripes. In this population, you do not see any variation in stripe number within each sex. Do you need to perform a statistical test? A. Yes, because they are comparing two conditions (males vs. females). B. Yes, because it is good scientific practice to perform statistical tests after data collection. C. No, because the number of individuals sampled is high (n=1000). **D. No, because there is no variation in these populations with respect to number of stripes.**
	Interpreting p-values generated by statistical tests	Fred is doing an experiment to test if water temperature affects zebrafish lifespan. He grows 2 tanks of zebrafish, one at 28ºC, and one at 25ºC. He analyzes ***<u>300</u>*** fish per condition. His data is shown below: 

		Fred performs a t-test to determine whether there is a significant difference between the two treatments, and calculates a p-value of 0.8. What can Fred confidently conclude about his data? A. This experiment is inconclusive, because the p-value is greater than 0.05, so his data is not informative. B. This experiment is inconclusive, because his sample size is too small to yield informative results. C. Water temperature does affect lifespan, because the range of lifespan is larger in fish grown at 25ºC. **D. Water temperature doesn't affect lifespan, because there is no significant difference in lifespan of fish grown at the two temperatures.**

From: Hicks, J., Dewey, J., Brandvain, Y., & Schuchardt, A. (2020). Development of the biological variation in experimental design and analysis (BioVEDA) assessment. *PLOS One, 15*(7).