TEACHER'S TOOLKIT

An inquiry-based activity to show the importance of sample size and random sampling

BY CHRISTOPHER DOBSON, M. MEGAN WOLLER-SKAR, AND JARED GREEN

This article describes our inquiry-based activity that teaches students the importance of random sampling, a critical component of many ecology investigations. Planning and Carrying Out Investigations is a key science and engineering practice of the *Next Generation Science Standards* (*NGSS*) (NGSS Lead States 2013), which is based on *A Framework for K–12 Science Education* (National Research Council 2012). Two of the specific components of the practice that students are expected to master include (1) evaluating and revising experimental design to produce data; and (2) evaluating the accuracy of various methods for data collection. Our activity allows students to discover the effect of sample size and sampling bias, essential considerations in improving the accuracy of data and experimental design.

Our random sampling activity can easily be used in conjunction with ecology-related lessons or student investigations of a natural area on a topic of their choosing. For example, our students have investigated fungal abundance on fallen logs, substrate preferences of pond snails, and water strider behavior in small streams. Figure 1 shows two middle school students using random sampling to estimate the abundance of stream algae in a study. Students are using a 10 cm × 10 cm sample plot made of bending straws with the ends slipped into each other. The sample plot fits a total of 200 times in the 1 m × 2 m section of stream delineated with string. To provide your students the skills necessary for carrying out a similar procedure, we suggest conducting the described classroom activity early on to help students learn the importance of random sampling in the planning stages of their investigations. We conduct our random

FIGURE 1: Middle school students use random sampling to estimate the abundance of stream algae

ALL PHOTOS COURTESY OF THE AUTHORS

FIGURE 2: Random number table

01128	32477	55013	73143	41204
03980	87195	14015	21283	04711
27429	14610	03363	12645	59051
40261	33266	96167	60022	20461
86382	75291	37134	71020	15457

sampling activity during the time period when students are working on their experimental design and before they begin data collection.

Scientists use sampling to estimate various parameters of a population, such as population size, when counting or measuring the entire population is impractical. Sampling seldom results in a completely accurate estimate of the population, but a larger sample size reduces variation and increases the likelihood that an estimate is reasonably accurate. Suppose students wanted to estimate the number of aphids on the leaves of a rose bush with 100 leaves. They could count the aphids on five leaves (samples), calculate the mean, and multiply by 20 (5 × 20 = 100) to estimate the total number of aphids on the entire bush. Counting the aphids on 10 leaves and multiplying the mean by 10 (10 × 10 = 100), however, would likely produce a more accurate estimate.

Investigators can be biased when selecting samples to measure or count, choosing the largest, healthiest, or most noticeable. For example, students estimating the population size of dandelions on the school grounds might choose to select samples in regions where thick patches of the plant are growing. This would result in an overestimation of the number of dandelions. Clever students might choose to sample areas with very few dandelions, reducing the amount of time required to count them. This would result in an underestimate of population size. To obtain a more accurate estimate of population size, samples must occur in areas with small and large numbers of dandelions. Random sampling reduces bias by ensuring that all individuals or areas under investigation have an equal chance of being sampled.

Random number generators, which are available for free online (see Resources), can facilitate the random selection of areas to be sampled. They produce a list of random numbers that can be used to select individuals or areas to sample. The generators generate the number of sets of random numbers the user requests, after specifying simple criteria such as: (1) how many numbers per set; (2) the range of acceptable numbers (e.g., 1–50); and (3) whether numbers should remain unique, appearing only once.

Random number tables (Figure 2) present an easy-to-use alternative to random number generators. These are also available for free online, for example, from Million Random Digits (see Resources), or they can be found in the appendix of many introductory statistics texts. Random number tables can be printed or

CONTENT AREA

Population sampling

GRADE LEVEL

7

BIG IDEA/UNIT

Students learn the importance of sample size and random sampling for improving the accuracy of data and experimental design

ESSENTIAL PRE-EXISTING KNOWLEDGE

None

TIME REQUIRED

One hour

COST

Free if using materials you already possess; $15 if you purchase plastic centimeter cubes and flexible straws

FIGURE 3: Distribution of Kisses across table top, where each dot represents a single Hershey's Kiss

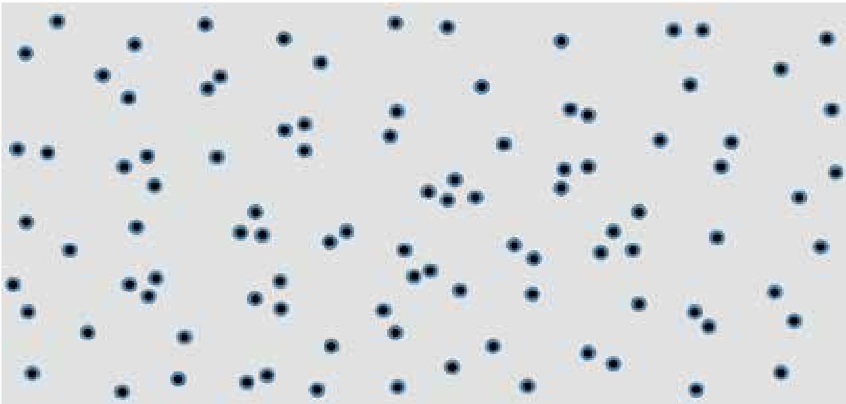
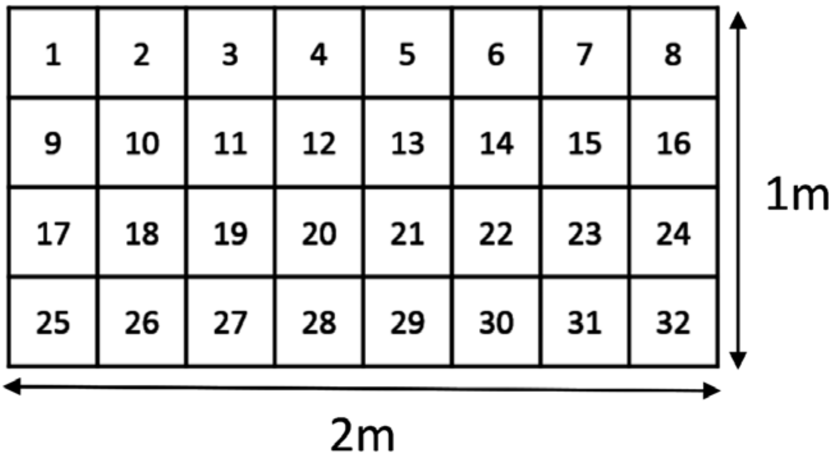


FIGURE 4: Sample plot fits 32 times across table top



photocopied and provided to students as handouts in class. To use a random number table, students arbitrarily select values by closing their eyes and pointing to a group of numbers. Then they read and compile a list of the two-digit numbers that occur in a given row, proceeding to the right-across columns. Any numbers outside the acceptable range are disregarded. Suppose a student begins selecting values in Figure 2 at 87,195, and the range of acceptable numbers was 1–50. Reading across the table, the 87 would be disregarded, and 19 would be the first acceptable number. Continuing to the right from 87,195 to 14,015, 51 would be disregarded, and 40 and 15 would join the list of acceptable numbers. If required, students continue reading into the next row that starts with 27,429, and so on until the desired quantity of random numbers is obtained.

Activity setup

There are no particular safety issues with this activity. We use Hershey's Kisses because they engage and entice students, leading to biased sampling, but we have the luxury of a food-safe classroom. Food should not be consumed in a science lab or classroom, so you can substitute Kisses with any small, immobile objects. For example, plastic centimeter cubes would work well.

We spread out 100 Kisses across the tops of three 1 m × 2 m tables (Figure 3). We position Kisses on the tables purposefully to illustrate the importance of random sampling (Figure 3). Throughout the middle of a table, we create clumps of three or four Kisses, with an occasional individual Kiss or clump of two. In the corners and along the edges of a table, only individual Kisses and clumps of two occur. We use square pieces of Plexiglas (0.25 m × 0.25 m) as a sample plot that students overlay on the Kisses to estimate population size. The sample plots fit a total of 32 times, four rows and eight columns, across an entire tabletop (Figure 4). To estimate population size, students multiply the kisses covered by the plot by 32. For example, a sample plot overlaying four Kisses would estimate the population size to be 128 (4 Kisses × 32).

If you lack tables of an appropriate size, conduct the activity on the floor of any available space. Sample plots can be constructed to represent whatever fraction of the sampling area you choose (1/32 in our example). Colored tape can be used to create a sampling area (table top) of any dimensions you like on the floor. Flexible plastic straws or pipe cleaners can be fastened together in a square shape to make a sample plot.

Biased sampling

We divide students into three groups and have them gather around one of three assigned tables. We do this to place a manageable number of students at each table (eight). More tables can be used, but this will increase setup time. We tell students they are going to estimate the population size at their table by laying the Plexiglas sample plot over some Kisses and multiplying by 32, since it is 1/32 of the table top. We remind students that actually counting the Kisses would defeat the purpose of the activity. We tell them to lay the sample plot down wherever they like.

Students tend to position the sample plots in a manner that includes as many Kisses as possible. For example, they often place them somewhere in the middle of a table, oriented strategically to include two or more clumps of Kisses; students typically do not place sampling plots in the corners or along the edges of the table, which would result in a smaller estimate of the population. This results in an overestimation of population size. In an example of biased sampling, the sample plot in Figure 5 has been positioned to include six Kisses. Next, we ask three of the students at each table to take a turn (sample size = 3) using the sample plot to estimate the population size. Having more students at each table take a turn simply increases sample size and the duration of the activity. We provide calculators and tell students to multiply the number of Kisses in

FIGURE 5: Biased placement of sample plot

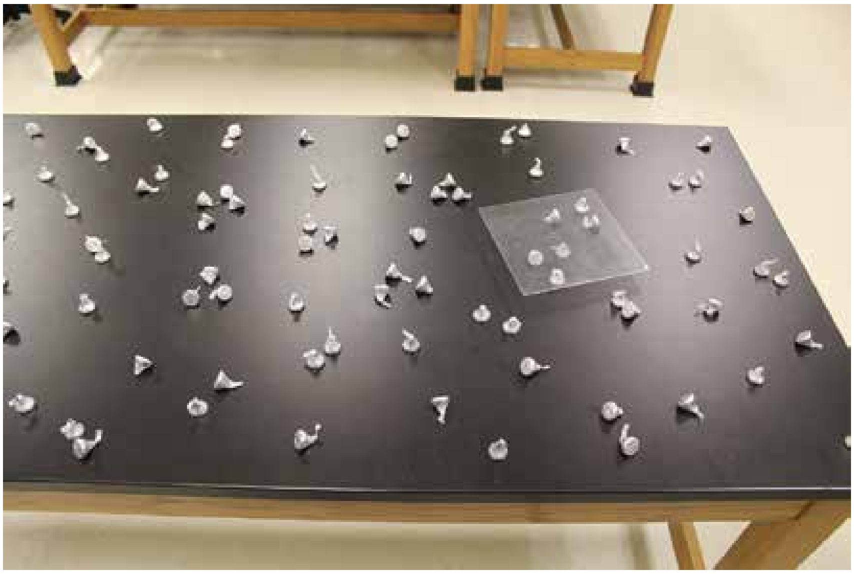
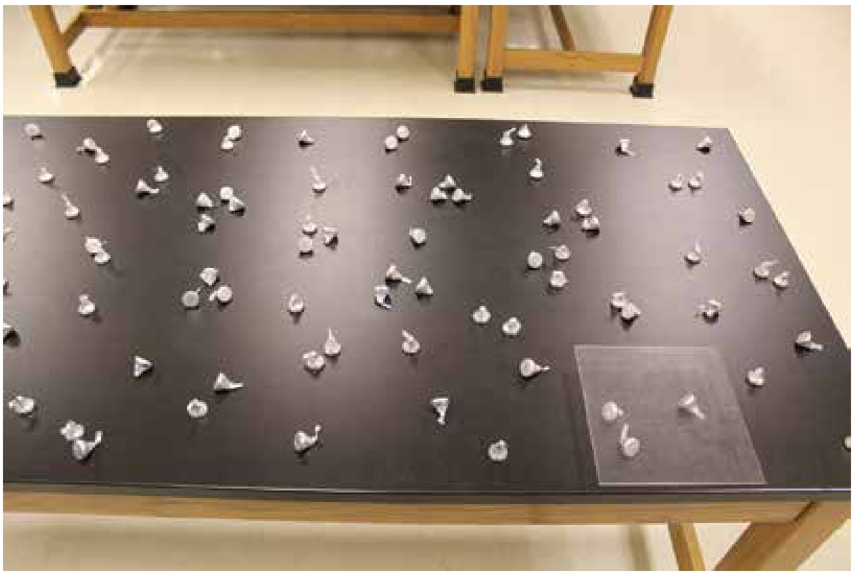


FIGURE 6: Random placement of sample plot



each sample plot by 32, as we record their findings on the board.

Random sampling

We introduce students to the concept of random sampling, explaining that their interest in the Kisses may have led to sampling bias and an overestimate of the actual population size. As a class we use a random number generator from Research Randomizer or Stat Trek (see Resources) to select samples for each group. We generate a set of three random numbers for each group. We draw Figure 4 on the board to show where each randomly selected sample is located on the tabletop. Then students repeat the previous procedure of laying the sample plot down and multiplying the number of Kisses they find by 32, as we record their data on the board. This random method often results in sample plots being placed in corners or along the edges of the table, locations students seldom choose to sample on their own. Figure 6 shows a randomly selected sample plot that occurs along the edge of the table (random number = 31) and includes only three Kisses. Even if a randomly selected sample plot occurs in the middle of the table over a larger clump of Kisses, it is not strategically oriented to maximize the number of Kisses to be counted, as in Figure 5.

Figure 7 contains the actual data from one of our classes as it appeared on the board. The estimates produced by each group in biased and random sampling are shown in columns with their means presented below. An average of the three group means is calculated as a "grand mean" to represent the overall mean estimates from biased and random sampling. We have no formal assessment for this activity. Rather, we use student responses to our questions to both drive discussions and gauge student understanding. We ask students why the estimates tend to be larger with the sample plots they "chose" relative to those that were randomly selected. They readily admit that they were tempted to put the sample plot over larger numbers of Kisses. We ask them how they think this may have influenced the estimates, and they generally acknowledge that they likely overestimated the population size. In our ecology unit, we look for the incorporation of random sampling by students in their investigations as a summative assessment.

We tell students to look at the actual tables with Kisses and inform them that there are the same number on each table. We ask them to guess how may Kisses are in each population. Most students guess a number near the correct answer of 100. When we confirm that each population consists of 100 Kisses, students are impressed with the ability of random sampling to produce a close estimate (grand mean = 106). They are also surprised by how much biased sampling overestimated the population size (grand mean = 199), an estimate double the actual number.

Next, we help students discover the effect of sample size by looking at the estimates from random sampling (Figure 7). We point out that the first estimates of Group 1 and 2 (64) are very different from the first estimate of Group 3 (192), and that neither estimate is very close to the

FIGURE 7: Data from class written on the board (actual number of Kisses = 100)

Biased sampling				**Random sampling**			
	Group 1	Group 2	Group 3		Group 1	Group 2	Group 3
	192	224	224		64	64	192
	96	256	192		64	96	96
	181	96	256		96	96	192
Mean	181	192	224	Mean	74	84	160
Grand mean = 199				Grand mean = 106			

accurate value of 100. Then we point out that the means for each group (74, 84, and 160) get closer to the actual population size, but that they are still not as close as the grand mean (106). We ask students to explain why. They generally say something like, "depending on where the random numbers had us place the sample plot, we got just a few, or a lot of Kisses, but that doing it many times gets us closer to the actual number." In this case, nine total estimates (sample size = 9), three from each group, resulted in a grand mean estimate of 106, a value very close to the actual number of 100. We make sure students understand that sampling will rarely produce a completely accurate estimate, but that random sampling and a large sample size increase its accuracy.

Conclusion

As described earlier, two specific components of Planning and Carrying Out Investigations require students to evaluate both experimental design and the accuracy of methods for data collection. This activity helps students learn the importance of random sampling and having a large sample size, knowledge critical to evaluating experimental design and the accuracy of methods for data collection.

The use of random sampling is a seventh-grade *Common Core* math standard, and Kisses in the Classroom is ideally suited to introduce students to the concept of random sampling prior to conducting ecology investigations in the field. We hope you will incorpore our random sampling activity in your curriculum to explore the importance of sample size and sampling bias, and to help improve the accuracy of data and the quality of your experimental design. ●

ACKNOWLEDGMENTS

We would like to thank Grand Valley students Alyssa Dehn, Jenna Dominique, and Shelby Fitzpatrick for their design of the stream algae investigation, as well as their photographic contribution, in Figure 1. We would also like to thank Stephen Rybczynski, an associate professor in the Biology Department at Grand Valley, for his photographic contributions in Figures 5 and 6.

REFERENCES

National Governors Association Center for Best Practices and Council of Chief State School Officers (NGAC and CCSSO). 2010. *Common core state standards*. Washington, DC: NGAC and CCSSO.

National Research Council (NRC). 2012. *A framework for K–12 science education: Practices, crosscutting concepts, and core ideas*. Washington, DC: National Academies Press.

NGSS Lead States. 2013. *Next Generation Science Standards: For states, by states*. Washington DC: National Academies Press. *www.nextgenscience.org/next-generation-science-standards*.

Connecting to the *Common Core State Standards* (NGAC and CCSSO 2010)

Mathematics

CCSS.Math.Content.7.SP.A.1. Understand that statistics can be used to gain information about a population by examining a sample of the population; generalizations about a population from a sample are valid only if the sample is representative of that population. Understand that random sampling tends to produce representative samples and support valid inferences.

RESOURCES

Ambrose, H., K. Ambrose, D. Emlen, and K. Bright. 2007. *A handbook of biological investigation*. 7th ed. Winston-Salem, NC: Hunter Textbooks.

Million Random Digits—*www.rand.org/pubs/monograph_reports/MR1418.html*

Research Randomizer—*www.randomizer.org*

Stat Trek—*http://stattrek.com*

Christopher Dobson (*dobsonc@gvsu.edu*) is a professor and **M. Megan Woller-Skar** (*wollerm@gvsu.edu*) is an associate professor in the Biology Department at Grand Valley State University in Allendale, Michigan. **Jared Green** is a graduate of Grand Valley State University.