STATISTICAL LITERACY
SIMULATIONS WITH

Jeremy Strayer and Amber Matuszewski

Swimming with dolphins can certainly be fun, but is it also therapeutic for patients suffering from clinical depression? To investigate this possibility, researchers recruited 30 subjects aged 18–65 with a clinical diagnosis of mild to moderate depression (Antonioli and Reveley 2005). Subjects were required to discontinue use of any antidepressant drugs or psychotherapy four weeks prior to the experiment and throughout the experiment. These 30 subjects went to an island off the coast of Honduras, where they were randomly assigned to one of two treatment groups. Both groups engaged in the same amount of swimming and snorkeling each day, but one group did so in the presence of bottlenose dolphins and the other group (outdoor nature program) did not. At the end of two weeks, each subject's level of depression was evaluated, as it had been at the beginning of the study.

—Rossman (2008), p. 9

A six-phase instructional strategy can help students develop conceptual understanding of inferential hypothesis testing.

DOLPHINS

n our data-rich world, statistical literacy is highly valued by employers and educators alike (Bargagliotti 2014; NCTM 2000). Although the Common Core State Standards for Mathematics places heavy emphasis on statistical topics in the middle- and high-school grades (CCSSI 2010), the American Statistical Association's "Statistical Education of Teachers (SET)" (Franklin et al. 2015) acknowledges that many practicing and preservice teachers have not had sufficient preparation to facilitate students' development of statistical literacy.

In this article, we present a six-phase strategy (see **fig. 1**) that teachers can use to help students develop a conceptual understanding of inferential hypothesis testing through simulation and address the content standard "Use data from a randomized experiment to compare two treatments; use simulations to decide if differences between parameters are significant" (CCSS.MATH.CONTENT.HSS .IC.B.5, CCSSI 2010). As we discuss the strategy, we describe each phase in general, explain how we implemented the phase while teaching our students

A Six-Phase Structure for Simulation Lessons

1. Commitment to a position in a rich context
2. Statement of possible hypotheses
3. Statement of expected results assuming the null hypothesis is true
4. Revelation of study results
5. Simulation under the null hypothesis
6. Making a conclusion

Fig. 1 Teachers designing lessons can use stages to help students develop conceptual understanding.

the Dolphins lesson found in appendix 1 of the "SET" publication (Franklin et al. 2015), and show how the phase aligns with teacher actions in *Principles to Actions: Ensuring Mathematical Success for All* (NCTM 2014) (see **fig. 2**).

The Dolphins lesson, based on a study found in the *British Medical Journal* (Antonioli and Reveley 2005), is one of many lessons that teachers can use to help students develop informal inferential statistical reasoning using simulations (Rossman 2008). In this article, we include our questioning techniques and responses to common student misconceptions to assist readers as they imagine how they might use the six-phase lesson strategy to implement other inferential simulation lessons in their own classrooms with their own students.

Teacher Actions Advocated by *Principles to Actions* and Supported by This Lesson

1. Supporting students in exploring tasks without taking over student thinking (*Principles to Actions* [NCTM 2014], p. 24)
2. Allocating substantial instructional time for students to use, discuss, and make connections among representations (p. 29)
3. Engaging students in purposeful sharing of mathematical ideas, reasoning, and approaches, using varied representations (p. 35)
4. Ensuring progress toward mathematical goals by making explicit connections to student approaches and reasoning (p. 35)
5. Making certain to ask questions that go beyond gathering information to probing thinking and requiring explanation and justification (p. 41)
6. Using visual models to support students' understanding of general methods (p. 48)
7. Giving students time to struggle with tasks and asking questions that scaffold students' thinking without stepping in to do the work for them (p. 52)
8. Eliciting and gathering evidence of student understanding at strategic points during instruction (p. 56)
9. Making in-the-moment decisions on how to respond to students with questions and prompts that probe, scaffold, and extend (p. 56)

Fig. 2 These teacher actions are referenced throughout the article.

THE SIX PHASES AND THE DOLPHINS LESSON

1. Commitment to a Position in a Rich Context

In the first phase, we engage students' interest by asking them to commit to a position. We asked students whether they believed that swimming with dolphins could relieve depression in people. A show of hands revealed that one-fourth of the class thought that swimming with dolphins would help relieve depression and three-quarters believed that swimming with dolphins would have no effect. With students committed to a position, we shared the description of the Dolphins study (see p. 606).

Next, we revealed a portion of the study's results—that 13 of the 30 participants showed an improvement in depression level. With this incomplete information, we asked students to commit once again while discussing this question in small groups: "Of these 13 improvers, how many do you think were in the dolphin group?" During the ensuing whole-class discussion (teacher actions 1 and 3; see **fig. 2**), many students qualified their answers, saying, "Well, I believe swimming with dolphins helps, so maybe 9 or 10 out of the 13 are in the dolphin group" and "I don't think swimming with dolphins helps depression, so about 6 or 7 will be in the dolphin group—it should be an equal split."

2. Statement of Possible Hypotheses

In the second phase, students extend their understanding of the qualifying statements they made in the previous phase. These statements are hypotheses that must be made before considering reasonable values for results in a study. We asked students to discuss this question: "What are all the possible hypotheses that could be made for this study?" (teacher actions 5, 7, and 8).

Some students struggled with this question, so we prompted them (teacher action 9), saying, "This study had people swimming or not swimming with dolphins, and researchers measured changes in depression levels. What do you think the researchers were hoping to show?" Another prompt we used was, "Earlier in the lesson you said, 'Swimming with dolphins helps.' That is one possible hypothesis. What are others?"

With this scaffolding, students identified three possible hypotheses: swimming with dolphins reduces depression levels; swimming with dolphins increases depression levels; and swimming with dolphins has no effect on depression levels.

3. Statement of Expected Results Assuming That the Null Hypothesis Is True

In the third phase, we introduce the null hypothesis as a statement that says that there is no difference between the two groups. We also tell students that

the inferential reasoning process assumes that the null hypothesis is true and then investigates the likelihood of randomly obtaining the results in the actual study.

In the Dolphins lesson, the null said that swimming with dolphins had no effect on depression levels, so we began the inferential reasoning process by asking students to discuss this question: "How many of the 13 improvers would you expect to be in the treatment group if the null hypothesis were true?" We ensured that students had sufficient time to discuss this question (teacher action 7) because although statistics teachers find this question natural, it takes students time to become comfortable thinking hypothetically.

When groups reported out, there was wide agreement that "if we assume the null hypothesis is true (swimming with dolphins does not affect depression levels), we expect 6 or 7 improvers in the dolphin group."

4. Revelation of Study Results

Now for the big reveal: Phase 4 discloses the actual results of the study and elicits responses from students. During the Dolphins lesson, we told students that 10 of the 13 improvers were in the dolphin group. We asked students to discuss whether these results were reasonable if the null hypothesis were true (teacher actions 1, 3, and 8). One student shared, "This result does not necessarily mean swimming with dolphins improves depression levels because it could have just randomly happened." Other students said the result did not mean anything because "the sample size is too small." Still others commented, "This is a big difference! It proves swimming with dolphins cures depression!"

Students needed to express their thoughts about the study's results so that we could build on those thoughts during the next phase (teacher action 4). We responded by asking, "How likely is it for these results to have randomly happened? In a sample of 30 participants, how likely is it for 10 of 13 improvers to randomly fall into the dolphin group if we assume that swimming with dolphins doesn't help? Let's simulate to find out!" (teacher action 5).

5. Simulation under the Null Hypothesis

We have found that although students may struggle with setting up a simulation, they better understand the purpose of the simulation when they determine their own simulation method (teacher action 1). In the Dolphins lesson, we provided different simulation materials (e.g., decks of cards, colored chips, and slips of paper) and asked students to use them to simulate the process of doing the Dolphins experiment assuming that the null hypothesis is true. We assisted students in planning
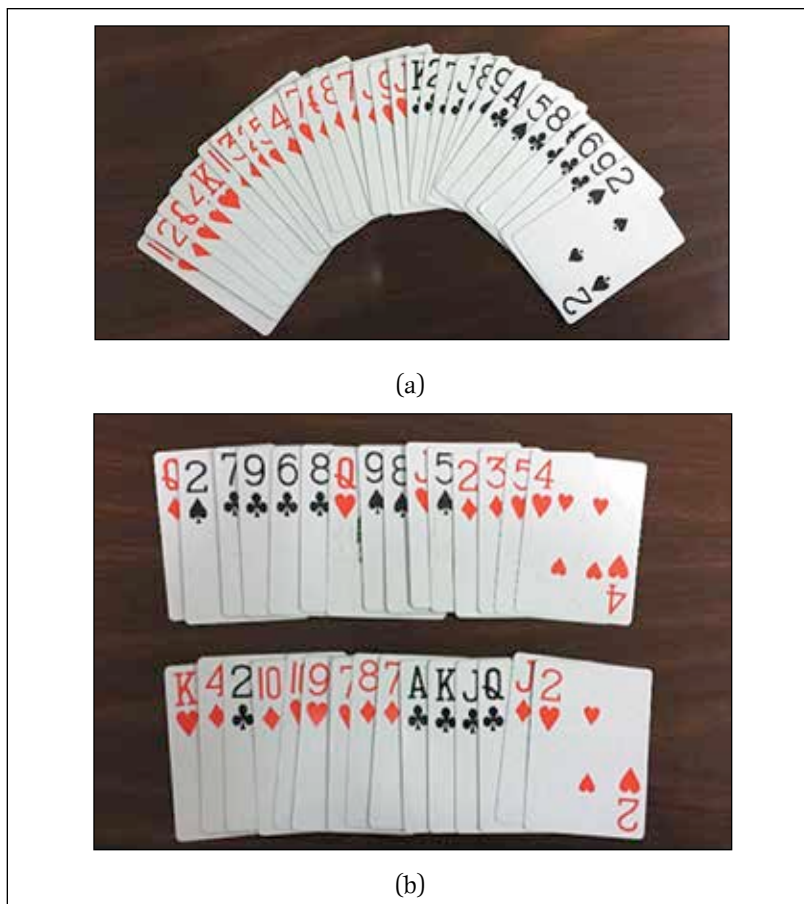

Fig. 3 Thirty cards represent the participants in the study, with improvers represented as black cards (**a**). Fifteen cards were randomly dealt to both the dolphin group (top) and nondolphin group (bottom) (**b**).

their simulations by asking, "How will you represent each participant? How will you represent the improvers (remember, there can be only 13)?" and "How will you randomly assign to the dolphin and nondolphin groups?" After a few minutes, students were ready to accurately simulate the study assuming that the null hypothesis was true.

Next, we asked students what information they should record as they simulate. To avoid taking over student thinking (teacher action 1), we had students discuss this question: "What outcome is of interest in this study, and how might we record it as a single measure?" Several students suggested that we record the difference between the numbers of improvers in each group.

One popular simulation method among our students was to use 30 cards to represent the study's participants and assign 13 black cards as "improvers" and 17 red cards as "nonimprovers" (see **fig. 3a**). Students shuffled and dealt 15 cards to two piles representing the dolphin and nondolphin groups. This shuffling simulated that "randomness" alone accounted for the assignment of improvers and nonimprovers to groups. Students counted how many improvers were in each group and recorded
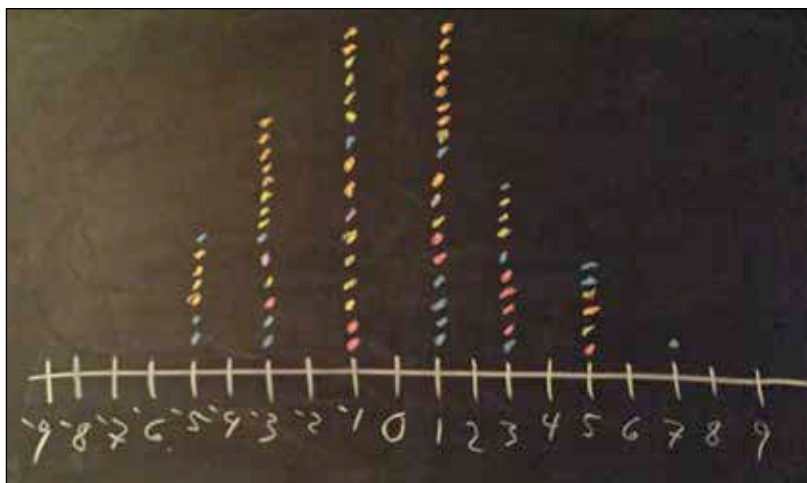
Fig. 4 Students create a dot plot on the board to display the difference in the number of improvers in their two groups in repeated simulations of the experiment.
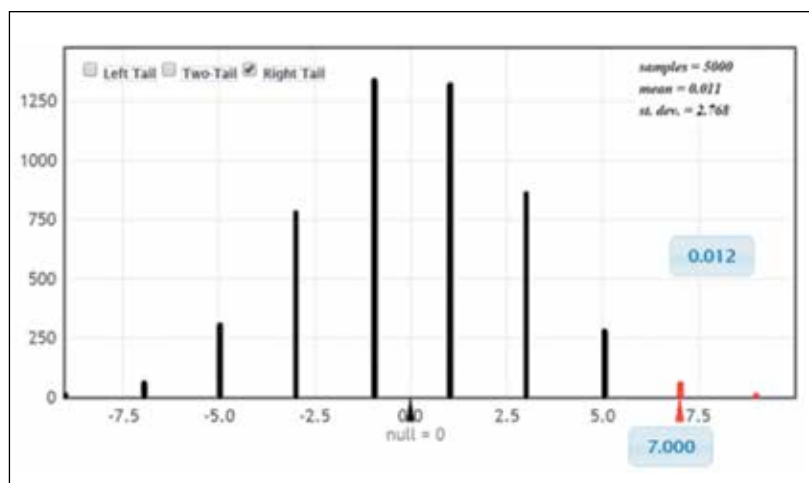

Fig. 5 A simulation of 5000 assignments of 13 improvers and 17 nonimprovers to groups shows that the dolphin group has 10 or more improvers (a difference of 7 or more) only 1.2% of the time.

the difference. In **figure 3b**, the dolphin group was the top set of cards, so the difference in improvers between the groups was 8 – 5 = 3. Note that if there were more improvers in the nondolphin group, the difference would be negative.

We asked students to repeat this simulation three times and record differences in improvers on a shared dot plot on the board (see **fig. 4**). Only 1 of 72 simulations had a difference as extreme as 7. Then we asked students to discuss this question: "How might we use a simulation like this to obtain a good estimation of the likelihood of getting

> **These ideas are at the heart of inferential statistical reasoning, and we wanted our students to construct their own understanding.**

results as extreme as the actual study (a difference of 7 or more)?" Some students suggested repeating the simulation thousands of times and finding the percentage of the time 7 or more occurred. Following their recommendation, we used free online technology (e.g., http://lock5stat.com/statkey/ and http://www.rossmanchance.com/applets/) to perform 5000 repeated simulations (see **fig. 5**) and found that a difference of 7 or more occurred only 65 times—an empirical probability of 0.012, or 1.2%. This is the critical mathematical finding. When we assumed that the null hypothesis was true and repeatedly simulated this study, the simulation showed that the study's actual results were unlikely. There appears to be only a 1.2% chance that randomness alone accounts for the difference of 7 that the researchers found in the Dolphins study or a value more extreme than 7. This measure of "unlikeliness" (0.012 in our simulations) is called the *p*-value.

These ideas are at the heart of inferential statistical reasoning, and we wanted our students to construct their own understanding of these ideas for themselves (teacher action 1). The final phase of the lesson provided students with this opportunity.

6. Making a Conclusion

In the final phase, we ask students to make conclusions about the original study's results based on the repeated simulations in phase 5 and to summarize their conclusions using formal statistical language. In the Dolphins lesson, we asked student groups to reflect on the representation of the repeated simulations (see **fig. 5**) and discuss the question, "What conclusions do these simulations allow us to make about the results of the original study?" (teacher actions 2, 3, 6, and 8). When groups reported out, some contended, "This simulation doesn't mean anything because our sample is too small." Others said that the researchers should repeat their study again before they could confidently draw conclusions. Still others believed that "swimming with dolphins does relieve depression" because the results (10 of the 13 improvers in the dolphin group) are not likely to have "just randomly happened."

To scaffold students' developing understanding, we asked student groups to discuss and respond to these objections (teacher actions 1, 4, and 9). During the whole-class discussion, many students relinquished the "small sample size" objection when one student explained, "All of our simulations used a sample size of 30. The difference of 7 was surprising precisely with a sample of 30!" Other students relinquished their "repeat the experiment" objection when one student shared, "We already repeated the experiment with our simulations—5000 times! The probability of getting a dif-

ference of 7 was very low based on the null hypothesis, so the null is probably not true." Some students pushed back against this statement: "Still, the researchers randomly could have gotten a difference of 7." This discussion helped students understand that inferential statistical decision making is based on probabilities, and although conclusions are not 100% certain, statisticians make informed decisions based on the collected data.

Finally, we provided students with definitions of terms—*null hypothesis*, *probability*, *p-value*, and *statistically significant*—and asked them to summarize their conclusions using these terms. Many students wrote summaries similar to this student's: "Believing the null hypothesis is true, the probability of the results being random are extremely low. Since the *p*-value is so small (0.012), the results are statistically significant. This means the null hypothesis is probably not true. Therefore, the dolphin treatment is a viable treatment for depression."

PEDAGOGICAL STRATEGY AND VISION

In far too many statistics classrooms, students memorize the procedural steps for hypothesis testing, find a *p*-value using technology, and make conclusions—all while barely understanding the concept of inferential statistics. Our six-phase strategy offers a different way. When we conduct several six-phase lessons with the same class of students, they come to deeply understand that the inferential data analysis process assumes the null hypothesis and uses probability to reason; what *p*-value is and how to talk about it; and the importance of context when making conclusions. Students need multiple opportunities to complete simulation lessons to develop these conceptual understandings. We encourage teachers to provide these opportunities while embracing the pedagogical vision set by *Principles to Actions* (2014) so that students can eventually develop strong procedural fluency for solving hypothesis test problems based on solid understanding of inferential statistical concepts (NCTM 2014).

REFERENCES

Antonioli, Christian, and Michael A. Reveley. 2005. "Randomised Controlled Trial of Animal Facilitated Therapy with Dolphins in the Treatment of Depression." *British Medical Journal* 331: 1231.

Bargagliotti, Anna E. 2014. "Statistics: The New 'It' Common-Core Subject." *Education Week.* January 29.

Common Core State Standards Initiative (CCSSI). 2010. Common Core State Standards for Mathematics. Washington, DC: National Governors Association Center for Best Practices and the Council of Chief State School Officers. http://www.corestandards .org/wp-content/uploads/Math_Standards.pdf

Franklin, Christine A., Anna E. Bargagliotti, Catherine A. Case, Gary D. Kader, Richard L. Scheaffer, and Denise A. Spangler. 2015. "The Statistical Education of Teachers." American Statistical Association. www.amstat.org/education/SET/

National Council of Teachers of Mathematics (NCTM). 2000. *Principles and Standards for School Mathematics.* Reston, VA: NCTM.

———. 2014. *Principles to Actions: Ensuring Mathematical Success for All.* Reston, VA: NCTM.
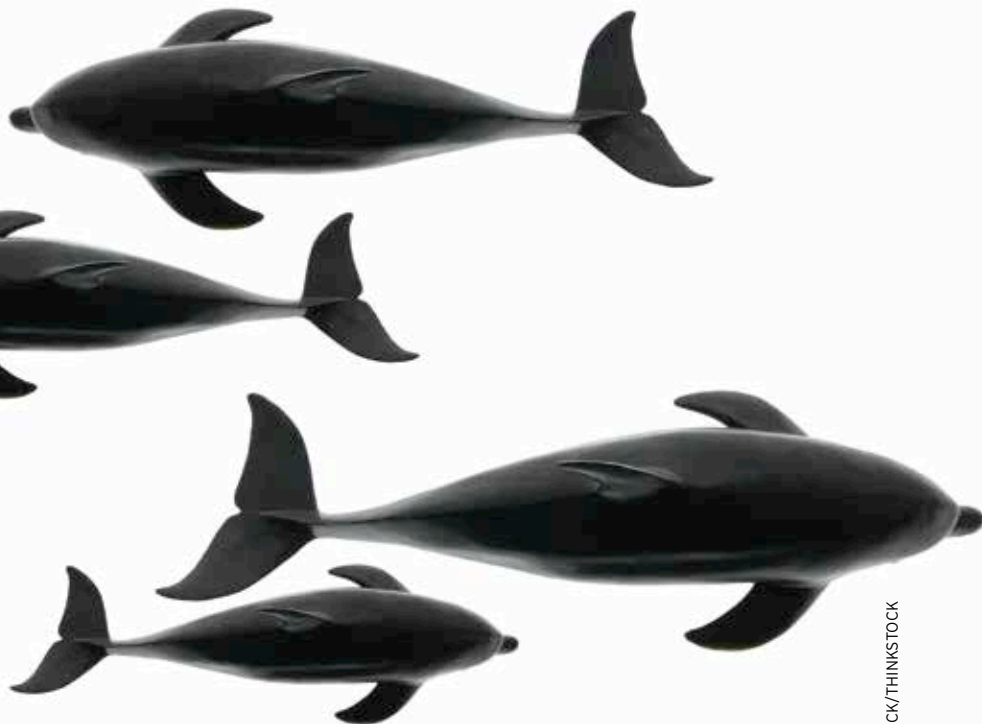
Rossman, Allan J. 2008. "Reasoning about Informal Statistical Inference: One Statistician's View." *Statistics Education Research Journal* 7 (2): 5–19.

JEREMY STRAYER, jeremy.strayer@ mtsu.edu, teaches in the Department of Mathematical Sciences at Middle Tennessee State University in Murfreesboro. He is interested in creating professional development opportunities to support K–16 teachers working to implement standards-based teaching practices.

AMBER MATUSZEWSKI, matuszewskia@rcschools .net, is the mathematics department chair at Siegel High School in Murfreesboro, Tennessee, where she teaches precalculus and AP Statistics. She is currently pursuing her PhD in mathematics education at Middle Tennessee State University in Murfreesboro.

SUNSTOCK/THINKSTOCK