

# Building Mathematical Models and Biological Insight in an Introductory Biology Course

A.E. Weisstein\*

Department of Biology, Truman State University, Kirksville, Missouri 63501, USA

**Abstract.** A growing body of literature testifies to the importance of quantitative reasoning skills in the 21st-century biology curriculum, and to the learning benefits associated with active pedagogies. The process of modeling a biological system provides an approach that integrates mathematical skills and higher-order thinking with existing course content knowledge. We describe a general strategy for teaching model-building in an introductory biology course, using the example of a model of an infectious disease outbreak. Preliminary assessment data suggest that working through the formal process of model construction may help students develop their scientific reasoning and communication skills.

**Key words:** mathematical modeling, education, mathematical biology, epidemiology

**AMS subject classification:** 92B01, 9701

## 1. Introduction

Many recent reports on the undergraduate biology curriculum have advocated a much stronger mathematical foundation for life science majors. For example, the *BIO2010* report recommends “building a strong foundation in mathematics, physical, and information sciences to prepare students for research that is increasingly interdisciplinary in nature” [17]. To a great extent, this mathematical foundation will also serve those students in life science courses who do not pursue careers in research. For example, national educational programs have called for teaching nonmajors STEM courses through large, open-ended problems that engage students in important questions of public policy [19]. Understanding such questions and evaluating potential courses of action requires that

---

\*E-mail: weisstae@truman.edu

students be able to interpret quantitative data and relationships [4]. Strong quantitative skills are thus crucial not only for future biomedical researchers, but for all future citizens of a society whose health and prosperity depend substantially on ongoing biological research.

The past 40 years have seen the development of new, active pedagogies such as case-based learning, inquiry-based learning, and workshop courses. These active teaching strategies seek to increase student engagement, content retention, and knowledge transfer to new experiences. Many studies have found these pedagogies to substantially improve students' retention and conceptual understanding of course material, as well as their ability to apply their knowledge to new situations [1, 2, 6, 13, 14, 20]. Several studies have also reported increased student understanding of the tentative nature of all science and more thoughtful evaluation of their own learning process [22].

We propose interweaving these two strands in biology education—mathematical biology and active learning—by having first-year biology students develop and analyze mathematical models of biological systems. This activity emphasizes the processes of identifying key biological assumptions, translating those assumptions into simple discrete-time recursion equations, and explaining the specific mechanisms by which changes in each parameter affect the simulated outcome. Through this modeling activity, we seek to equip students with a powerful and broadly applicable tool for approaching scientific questions, and to empower students to unpack mathematical relationships into explicit hypotheses that are subject to critical evaluation. More fundamentally, however, we intend this modeling activity to help students *build a more mature understanding of scientific processes*—an understanding that embraces stochasticity and competing processes, spans multiple levels of biological organization, and yields quantitatively testable predictions.

## 2. Course Structure

We use this activity in BIOL107, the first in a two-semester sequence of introductory biology courses aimed at science majors (including majors in allied health disciplines). Course enrollment is approximately evenly divided between first- and second-year students, and between biology and non-biology majors, with approximately 10% of the class consisting of upperclassmen and/or non-science majors. The course is taught as three 50-minute lectures and one 170-minute lab per week, all facilitated by the same instructor. Each lab section contains 20–26 students, and each lecture class combines students from 2–3 lab sections. In lab, two undergraduate teaching assistants help the instructor answer student questions and monitor each group's progress. Most BIOL 107 students have a mathematical background through calculus or pre-calculus, but few if any have prior experience using mathematical models.

Building and analyzing even a simple mathematical model is a time-intensive procedure. We therefore run the activity during lab to allow students enough time to develop a finished product while also building a conceptual understanding of the modeling process (approximately two hours total). In courses with a longer lab period, students could begin extending their model to specific infectious diseases (see below) or discuss a pre-assigned paper on epidemiology.

Most of the activity is completed in self-selected pairs of students. This arrangement provides ample opportunities for peer instruction, particularly when one student in the pair has substantially

more *Excel* experience than the other. At the same time, having only two students per group allows each student a reasonable amount of time at the computer keyboard and helps keep all students engaged. We also encourage students to master the entire process of modeling by assigning separate write-ups and by noting that their next lecture exam will include a brief modeling exercise.

Although the activity's main biological topic—the epidemiology of infectious diseases—is generally of great interest to students of all levels, it is not substantially addressed in most introductory textbooks (e.g., [5, 18], but see [8]). To integrate the activity into the course schedule, we have run a variant of the module in which students model the kinetics of an enzyme-catalyzed reaction rather than epidemiology (lesson plan available on request). This variant activity was performed during the fifth week of a 15-week semester, as part of the course unit on protein structure and function. By contrast, the SIR activity described here would be most appropriate either as a stand-alone exercise or in a redesigned introductory course that begins with population- and species-level phenomena and works its way down to molecular processes.

### 3. Introduction to SIR Models

The SIR model is a simple three-compartment model for an epidemic of infectious disease (see Figure 1). This model was introduced in 1927 by [12]. An SIR model divides a host population into three groups: Susceptible ( $S$ ), Infected ( $I$ ), and Recovered ( $R$ ). Movement of hosts between the model's three compartments reflects the progress of the epidemic.

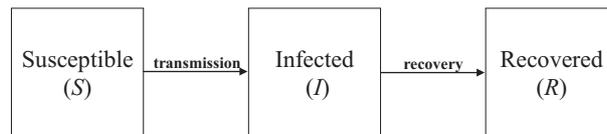


Figure 1: A box-and-arrow diagram of the Kermack/McKendrick SIR model. Boxes represent population compartments, while arrows represent movement of individuals between compartments.

The basic SIR model makes many simplifying assumptions, including:

- The time course of the epidemic is short enough that natural birth and death of hosts can be ignored.
- The mortality from the disease is effectively zero.
- The disease is transmitted directly between hosts, rather than by insect vectors, contaminated surfaces or water supplies, etc.
- Hosts do not vary significantly in their susceptibility, and there are no significant differences in virulence or transmissibility between disease strains.
- Recovery confers permanent immunity.

- There is no age, geographic, or social structure to the host population.
- Transmission and recovery rates are constant over the epidemic (i.e., no interventions such as quarantines, new treatments, closures of public facilities).

Obviously, most real diseases violate several of these assumptions. Once we have built a simple model, our next step will be building a model that more accurately captures the features of a specific disease.

### 3.1. Graphical Models: Encapsulating Biological Knowledge

The first step in our modeling approach consists of building a simple box-and-arrow diagram (Figure 1) that explicitly describes the population's compartments and the processes by which individuals move among them. In constructing this graphical model, students will decide which features of the system must be included and which can be omitted for the sake of simplicity. This process can help students see that there are many different ways to model the same system, depending on the specific question under study, and that even deliberately oversimplified models can nonetheless yield important insights.

We begin with a chalkboard diagram of a single square labeled  $S$ , which represents all the individuals within the population who are currently susceptible to the disease. We then add a second square  $I$  that represents the individuals who are currently infected with the disease, and a third square  $R$  that represents the individuals who have recovered and are resistant to further infection. Next, we ask students to describe the system's short-term behavior: "In the first few days of the epidemic, what is likely to happen to the susceptible individuals? to the infected individuals?" Students quickly agree that some of the susceptible individuals will become infected, and some of the infected individuals will recover. Asked how to depict this behavior on the chalkboard diagram, students usually suggest adding arrows between the appropriate model compartments. Once we have done so, labeling the arrows "transmission" and "recovery" for the sake of clarity, our first version of the model is complete.

At this point, students often suggest refinements to the model (e.g., latency periods, individual hosts with differing levels of susceptibility). We emphasize that we are deliberately making the model as simple as possible for now, to focus on the process of model construction rather than on generating a specific product. However, we note that these suggestions represent important assumptions contained in our model and record them on the blackboard. Students are told that they will later have the chance to extend the model to incorporate some of these features.

### 3.2. Writing "Word Equations"

Now that students have built a visual model, we next write "word equations" that explicitly but qualitatively encapsulate the system's major processes. Our goal is to have students express their proto-mathematical understanding of a system before introducing the formal mathematics. By deferring the formal mathematical symbolism until the next step, students can focus on the reasoning

process behind modeling. In effect, they are doing the thinking part of math, just not the mechanics that intimidate so many students.

We begin this section with the following simple example to illustrate the concept of a word equation:

“Imagine a coin jar. Let’s say we want to know how many coins were in the jar at the start of the month, on September 1st. What information would you need to determine how many coins there will be on October 1st?”

Initially, students may approach this problem from many different angles, assuming that the instructor wants a complex answer relating to the number of people in the household, the relative frequencies of different coins, and so on. However, the coin jar metaphor usually calls up ideas about adding and removing coins. With a little encouragement and an emphasis on the most fundamental aspects of the problem, students soon agree that they need three pieces of information: the number of coins on Sept. 1st, the number of coins removed during September, and the number of coins added during the same period. We write these terms on the board as follows (including the blank spaces):

$$\begin{array}{ccccccc} \# \text{ coins in jar} & & \# \text{ coins in jar} & & \# \text{ coins added} & & \# \text{ coins} \\ \text{on Oct. 1st} & = & \text{on Sept. 1st} & & & & \text{removed} \end{array}$$

Next, we ask students how the addition and removal of coins will influence the total number present in the jar on Oct. 1st. Although many students consider the answer rather obvious, the experience of being forced to explicitly consider this question comes in useful when they move on to more complicated examples. The final word equation for the coin jar example is:

$$\begin{array}{ccccccc} \# \text{ coins in jar} & = & \# \text{ coins in jar} & + & \# \text{ coins added} & - & \# \text{ coins} \\ \text{on Oct. 1st} & & \text{on Sept. 1st} & & & & \text{removed} \end{array}$$

Having worked through an accessible case, we then return to the SIR model. With Figure 1 and the coin jar example both still on the board, we ask students to write word equations for the number of (a) susceptible, (b) infected, and (c) recovered individuals on Day  $t + 1$ . Students work on this project in small groups for approximately five minutes while we write the left-hand side of Eq. (3.1) on the board and circulate to answer any questions. Once most groups have finished, we ask for volunteers to help us complete the equations, which are as follows:

$$\begin{array}{ccccccc} \# \text{ Susceptible} & = & \# \text{ Susceptible} & - & \# \text{ new} & & \\ \text{on Day } t + 1 & & \text{on Day } t & & \text{transmissions} & & \\ \\ \# \text{ Infected on} & = & \# \text{ Infected on} & + & \# \text{ new} & - & \# \text{ new} \\ \text{Day } t + 1 & & \text{Day } t & & \text{transmissions} & & \text{recoveries} \\ \\ \# \text{ Recovered} & = & \# \text{ Recovered} & + & \# \text{ new} & & \\ \text{on Day } t + 1 & & \text{on Day } t & & \text{recoveries} & & \end{array} \quad (3.1)$$

At this point, we take a few minutes to compare these equations to Figure 1. In particular, we note that the diagram shows that no processes can ever move infected or recovered individuals back to susceptibility, and that the only way for individuals to leave the susceptible category is by becoming infected. Accordingly, the word equation for susceptibles has no additive terms, and that the one subtractive term leads directly to the infected category. This discussion makes students comfortable with the notion that the word equations accurately describe their intuitive understanding of the system. To bring in one more mathematical concept at this stage, we guide students through the process of adding together the three individual word equations. With minimal prompting, most students correctly conclude that the left-hand side now represents the total number of all individuals on Day  $t+1$ , while the first terms on the right-hand side add up to the total number of individuals on Day  $t$ , and all the remaining terms cancel, leaving the result:

$$\begin{array}{l} \text{Total \# individuals} \\ \text{on Day } t + 1 \end{array} = \begin{array}{l} \text{Total \# individuals} \\ \text{on Day } t \end{array} \quad (3.2)$$

When asked to explain why this result should hold, students note that transmission and recovery simply move individuals from one population compartment to another. This leads to a discussion of the real-world circumstances under which this assumption is reasonable (e.g., an influenza outbreak on a college campus, but not a high-mortality or long-term epidemic such as Ebola or HIV). This discussion gives students an immediate payoff in terms of biological understanding and helps ground the lesson in familiar and relevant ideas before they embark on the next step of the modeling process.

### 3.3. Translating Word Equations into Formal Equations

By writing word equations, students have already built most of the framework for their mathematical model. The next step is to translate each verbal description (e.g., “# of new transmissions”) into explicitly mathematical terms. This is often the least intuitive step for students, and the step most likely to trigger disengagement in math-averse and math-phobic students. However, the supportive atmosphere of a small group and the instructor’s guidance through each step of the reasoning process can help minimize these difficulties. The ambitious but achievable goal of this step is a simple set of equations that students have written themselves, and for which they understand the biological meaning of each individual term.

We ease students into this process by starting with the simplest terms in Eqs. (3.1) and (3.2): the number of individuals in each population compartment on a given day. In our experience, most students immediately see that the number of susceptibles can be written as  $S$ , the number of infecteds as  $I$ , and the number of recovered as  $R$ . Moreover, they have little difficulty with using  $S_t$  and  $S_{t+1}$  to indicate the number of susceptibles on Days  $t$  and  $t + 1$ , respectively. We therefore quickly arrive at the following equations:

$$S_{t+1} = S_t - \begin{array}{l} \# \text{ new} \\ \text{transmissions} \end{array}$$

$$I_{t+1} = I_t + \begin{array}{l} \# \text{ new} \\ \text{transmissions} \end{array} - \begin{array}{l} \# \text{ new} \\ \text{recoveries} \end{array}$$

$$R_{t+1} = R_t + \begin{array}{l} \# \text{ new} \\ \text{recoveries} \end{array}$$

Next, we focus on the number of recoveries. To begin, we ask students what factors within the model will influence the number of individuals recovering on a particular day. As before, students may initially propose complex answers that invoke individual patient's overall age and health status, level of care received, and similar factors, all of which we record on the board. When we press students to relate their answers to specific parts of our simple model (Figure 1), they decide that the two key factors influencing recovery are the number of individuals currently infected and the overall recovery rate. We then ask students to work a simple numeric example: if there are 100 infected people at the start of Day  $t$ , and 8% of them recover that day, how many new recoveries have occurred? Even our most math-phobic students are usually very quick to respond correctly (8 new recoveries). Asked how they obtained their answer, students respond that they simply took 8% of 100, or  $8\% \times 100$ . We then ask students how to apply this same reasoning to the SIR problem, letting  $\beta$  represent the overall recovery rate. With the numeric example in front of them, students have little difficulty in seeing that the number of new recoveries occurring on Day  $t$  is given by  $\beta I_t$ .

We then turn to the number of transmissions. Students often view this case as exactly analogous to the number of recoveries, and predict that the number of transmissions will be equal to the number of susceptibles times the transmission rate  $\alpha$ . To test this hypothesis, we note that smallpox has an estimated transmission rate of 0.2–1.9% [16], and that routine smallpox vaccination in the United States ended in 1972 [3]. Students then discuss in small groups how many new cases of smallpox they expect to occur on campus tomorrow, and their reasoning process. Some students may know that smallpox has been eradicated, but even groups lacking this knowledge usually conclude that no new cases can arise (despite the large pool of unvaccinated susceptibles) due to the absence of any infected individuals on campus. Predicting the number of new transmissions thus requires knowledge of *three* factors: the overall transmission rate, the number of susceptibles, and the number of infecteds.

Rather than simply telling students how to combine these three factors, we use a simple example to demonstrate the underlying logic. Specifically, we ask students how many transmissions would occur in each of the following populations:

**Population A:** high transmission rate and many susceptibles, but zero infecteds.

**Population B:** high transmission rate and many infecteds, but zero susceptibles.

**Population C:** many susceptibles and infecteds, but a zero transmission rate.

As in the previous smallpox scenario, students recognize that no transmissions will occur in any of these populations. Therefore, we conclude that we need some mathematical operation that yields a zero result if any one of the three terms is zero. This phrasing usually elicits the suggestion to multiply the three terms together. Testing this idea on Populations A–C indicates that it behaves as desired, yielding the result that the number of new transmissions on Day  $t$  is given by  $\alpha S_t I_t$ .

Finally, student groups use the formulas we have derived to rewrite the word equations as true mathematical expressions (Eq. (3.3)). This process takes roughly 3–5 minutes. To reinforce their learning and detect any lingering confusion, we also ask students to explain in their own words the specific meaning of individual terms; for example, the term  $\beta I_t$  represents the number of infected individuals recovering from the disease on Day  $t$ . Once this is complete, we briefly pause to reflect on what we have accomplished so far: using only their prior biological knowledge and basic mathematical skills, students have successfully written a set of equations that describe an infectious disease outbreak over time. They can now convert the resulting model into a computer simulation.

$$\begin{aligned}S_{t+1} &= S_t - \alpha S_t I_t \\I_{t+1} &= I_t + \alpha S_t I_t - \beta I_t \\R_{t+1} &= R_t + \beta I_t\end{aligned}\tag{3.3}$$

### 3.4. Implementing the Model

Next, we implement the model in *Excel*. While there are many excellent modeling software packages (e.g., Berkeley-Madonna, Stella, MATLAB), we chose *Excel* for its ubiquity and familiarity to students. In addition, by displaying the results of each calculation as soon as students enter the formula, *Excel* drives home the point that the equations are not just some abstract representation that must be converted into a working mathematical model. Instead, the equations themselves *are* the model.

Using *Excel*'s built-in Name Box, students first assign informative names to model parameters, thereby making their equations easier to understand and proofread. They then enter column headers and initial values before typing the recursion equations for Day 1 of the simulation. We found that even students with prior Excel experience had difficulties with the process of entering formulas: a “cheat sheet” including templates for common formulas such as simple arithmetic and absolute vs. relative references may be helpful. Once the formulas are entered, a quick Copy-and-Paste (with one row for each day of the simulation’s time course) completes the initial model. The entire process of entering the equations into Excel takes an average of 10 minutes. To keep all students fully engaged, we encourage groups to move on to the next step as soon as they have entered their model (Figure 2).

Next, we reconvene as a class to demonstrate how to insert and fine-tune an *Excel* graph. Many students are used to simply printing off the very first graph they get, so we spend approximately 10 minutes as a class discussing appropriate format and content. The prompt “What specific changes could we make to improve this graph?” usually elicits ideas such as re-scaling axes, specifying

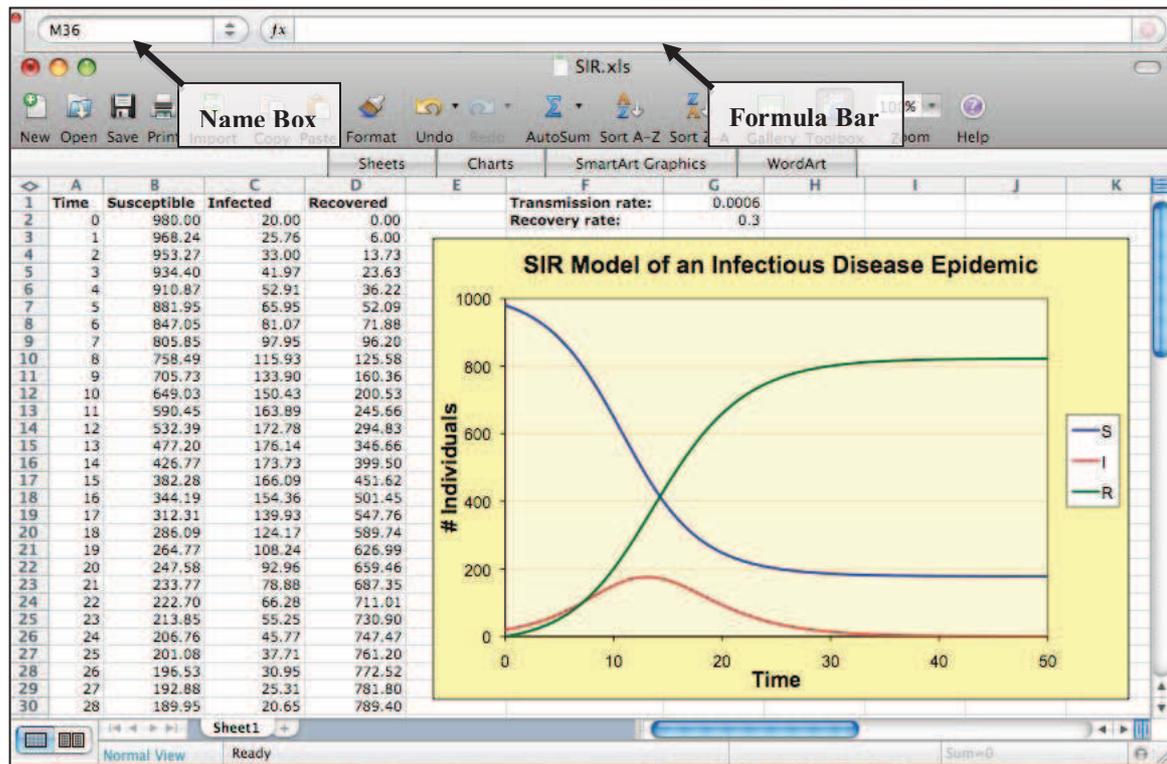


Figure 2: Screen shot of a working SIR model, showing the time course of an epidemic under typical parameter values. Controls with which students may be unfamiliar are labeled.

units, adding an informative title, and even choosing an intuitive color scheme (e.g., susceptibles = blue, infecteds = red, recovereds = green).

Finally, we ask students to explore the behavior of their new models. The lab handout includes a list of 6–8 questions that relate the model back to specific biological questions (see Table 1). Each group of students chooses one of these brief investigations. After 20 minutes, a spokesperson from each group gives a 2-minute informal presentation of their findings and of any new questions that arose.

### 3.5. Extending the Model

As a final group exercise, we ask students to consider how to extend the model. The simple SIR model, we point out, omits many common features of real diseases. Students pair up and choose a specific disease to model, then list 3–4 biological features of that disease that are *not* included in the basic model and propose a specific strategy for adding that feature to the model. For example, one pair of students might model the incubation period of measles by adding a fourth compartment  $E(t)$  during which patients are infected but not yet symptomatic or infectious. Another pair of students might model the effect of human genetic variation on resistance to malaria by splitting the

1. What effect does changing the disease's *transmission rate* have on the course of the epidemic? Clearly describe both the specific changes that occur in your graph and the biological meaning of those changes. Then do the same for the *recovery rate*.
2. How could you model a population in which a certain percentage of individuals are initially *vaccinated* against the disease? When you make this change in your model, what happens to the epidemic? Clearly describe both the specific changes that occur in your graph and the biological meaning of those changes.
3. Run the model several times without changing the parameters. How closely do the results of each run match each other? Explain why, clearly relating your answer to the equations underlying the model. Is this behavior realistic? If not, what *general type of changes* could you make to the equations to yield more realistic behavior?

Table 1: Examples of Initial Student Investigations with their *Excel* models.

Susceptible compartment into two subcompartments  $S_1$  (normally susceptible) and  $S_2$  (partially resistant). Students present the disease features they chose to model and their strategies for doing so. Classmates and the instructor then evaluate those strategies and suggest possible alternatives, underscoring that there are many different ways to model the same system.

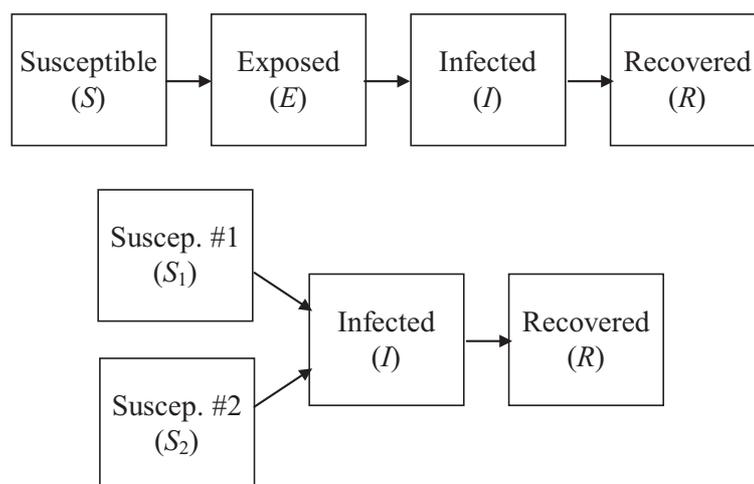


Figure 3: Box-and-arrow diagrams for two extended models (compare with Figure 1).

Whenever we have used this activity, whether in a formal course or as a one-hour guest lecture, students have consistently responded with great enthusiasm to the idea of modeling specific diseases. The ensuing conversations about which features are most important to include, and of how to model those features, provides a transition from a fairly structured exercise to an open-ended investigation with clearer relevance to real-world questions. We have also done additional follow-up in which three students asked to develop their models further as an independent studies project

out of class time. One student's project focused on long-term changes in HPV prevalence among adolescents in response to different vaccination schedules; the other two collaborated to model the spatial dynamics of a bioterrorism attack using Ebola virus. These students devised more complex models, reviewed the literature for relevant parameter values, and presented their results in a formal symposium. The exercise can thus be extended in many different directions depending on the effort that faculty and students wish to invest.

## 4. Learning Outcomes

Previous studies have identified a number of ways in which mathematical models contribute to understanding biological systems. These contributions include clarification of relationships between components, comparison of alternative conceptualizations, and identification of general principles [21]. In our work with SIR models, we noted four main learning outcomes: (i) improved equation literacy, (ii) greater conceptual and descriptive precision, (iii) formation of conceptual connections within and among disciplines, and (iv) more mature scientific judgment. We elaborate on these outcomes below.

Note that the original activity lacked any genuine assessment. Voluntary post-activity assessment was conducted approximately four months later, in the form of three open-ended prompts sent to all students in the course. Only three out of approximately 50 students responded; the results should therefore be considered as anecdotal rather than representative. Ideas for more meaningful assessment are discussed under "Future Directions."

*Equation literacy* has been defined by [11] as "the ability to translate from a holistic understanding of a situation into a quantitative equation-based description, and vice versa." This skill has traditionally been associated mainly with math and physics classes, but has obvious relevance to modern biology as well. We are not attempting to transform all biology students into professional modelers, but rather to transform the way they interact with equations. In our own experience, many first- and second-year courses treat equations (if at all) as "black boxes" into which one merely plugs numbers and calculates answers by rote [9]. By contrast, students noted that building their own equations made them reflect on the biological meaning of each variable and the specific interactions between them. One student explicitly compared the initial process of writing equations to translating into a foreign language, while commenting on how it helped her recognize some of her own misconceptions about the system. We consider this ability to critically analyze equations by extracting and examining their underlying biological assumptions to be a key outcome of the modeling process, and one with strong potential for transfer to other courses.

*Precision* in understanding and communicating complex concepts is a key skill that all incoming college students must learn, regardless of their major. Introductory-level students are often accustomed to a superficial level of understanding that masks fundamental misconceptions or gaps in their knowledge. For example, students may describe osmosis as "water wanting to spread out evenly across a barrier" and natural selection as "a species gradually getting better over time." Even if these students can give more precise answers when prompted, they may not fully grasp why their initial answers were inadequate. The process of building and implementing a mathe-

mathematical model continually confronts students with the question “What exactly does that mean?” For example, students may decide to add a latency period to their extended SIR model. As they write the model equations, however, they will need to determine what a “latency period” actually means. Symptoms precede infectiousness in some diseases, while the reverse is true for others. As a result, latency can refer to either (i) the period between infection and symptom onset or (ii) the period between infection and patient infectiousness.

In a survey of professional ecologists, respondents were asked to select the most important reason for using quantitative models. The two most common responses were “clarifies conceptualization of system structure” and “clarifies quantitative relationships. . . among system components” [15]. If practicing scientists find that building models helps them think more clearly about complex systems, surely our students will benefit from that same opportunity to refine their reasoning.

This exercise is also intended to shift students’ learning approach by spotlighting *connections within and between disciplines*. First-year students often view biology as a set of unrelated facts to be memorized. For example, over the course of a semester, students may learn how enzyme and substrate concentrations influence reaction rate, how concentration gradients and membrane permeability affect a cell’s osmotic balance, and how transmission and recovery rates determine the outcome of an epidemic. However, many students are likely to focus solely on the minute details of these seemingly unrelated processes, thereby missing the key insight that all three are simply instances of the same underlying phenomenon (i.e., movement of particles among different states). By developing model-based thinking early in the semester and explicitly framing subsequent lessons in terms of the rules that govern a system’s behavior, we are attempting to help students develop their ability to abstract from specific cases to more general principles, so that they can more easily transfer their knowledge from one system to another. For example, many first-year science majors have at least some familiarity with the principle that energy is conserved. Some may also be familiar with other conserved physical quantities (e.g., mass, charge, linear momentum). In working through the SIR modeling activity, we emphasize the applicability of conservation laws in an unfamiliar context; namely, that in a closed population with no birth, death, or migration, the total number of individuals summed across all categories remains constant (Eq. (3.2)). While this may seem obvious, it comes as a revelation to many students that conservation laws can be applied beyond the sometimes contrived setting of a physics problem. In extending their models to include births and/or deaths, students confront the idea of an open system, motivating a discussion that can lead to a broader classroom exploration of energy and matter flows into and out of individual cells, organisms, and entire ecosystems.

*Mature scientific judgment*—the ability to make reasonable, informed decisions that weigh a specific factor’s biological importance against computational or conceptual complexity—is among the most vital skills in conducting any open-ended research project. Pedagogies that seek to foster learning through a process of genuine scientific inquiry thus give students many opportunities to develop this skill through designing their own experiments, selecting appropriate modes of analysis, and defending their decisions and conclusions [10]. The SIR exercise directly addresses this issue by presenting a highly simplified model that nonetheless captures key behaviors of the epidemiological system (e.g., a vaccination threshold above which the epidemic cannot take hold). In their informal feedback, one student commented on how much she enjoyed drawing her own con-

clusions based on a model she had written herself, while another student noted that he had never realized how even a simple model could provide so much insight.

The activity's final step, in which students extend the model to a specific disease of their choice, is particularly effective at helping students begin to see the importance of prioritization and approximation in science. During the independent summer project, for example, the HPV model was built to include rates of sexual contact between teenagers of all age combinations. The student quickly found that precise data on such rates were virtually unobtainable, but realized that a single contact rate across all groups would be unrealistically simple. He therefore settled for using best-guess estimates that reflected assortative contact by age, accepting a potential loss of accuracy for the sake of arriving at a reasonable initial estimate. Similarly, and unsurprisingly given their project's security implications, the two students working on the Ebola model could not find any detailed data on the virus's transmission rate and latency period. Instead, they had to track down multiple sources in the public domain and then piece together reasonable estimates of these parameters.

This anecdotal evidence supports the idea that students develop scientific reasoning skills through practicing, not just passively learning about, science. If further study supports the hypothesis that open-ended modeling projects can help students develop as critical thinkers capable of exercising sound scientific judgment, that outcome alone may make such projects worthwhile in the classroom.

## 5. Taking the Next Step: Assessment

While our experiences with the modeling exercise have been highly positive, we acknowledge that they cannot be considered reliable without formal assessment of our four main learning objectives. Because we specifically want to measure students' ability to apply these skills to unfamiliar material, we anticipate some difficulties in preparing questions that are not biased by pre-existing differences in student knowledge of biological content. However, short essay prompts with coded rubrics [7, 22] may allow the desired resolution of students' critical thinking skills as they relate to modeling. Prototype questions for each objective are shown in Table 2, along with some of the main concepts being measured.

## 6. Conclusion: Model-Based Thinking across the Curriculum

Many first-year students may feel overwhelmed by the idea of building mathematical models and may not see any real point to doing so. We therefore propose beginning with a system of which students have some prior knowledge and that has obvious real-world implications. Epidemiology is an example of one such system, but many other systems (e.g. pharmacokinetics, community dynamics) could serve equally well. Each such activity empowers students to approach complex questions by translating vague initial descriptions into models that explicitly incorporate precise hypotheses about underlying processes.

Learning outcomes such as equation literacy and mature scientific judgment are easily lost if not subsequently reinforced. However, once students have gained some experience in modeling,

Objective	Question prompt	Key concepts
1. Equation literacy	The equation for a system's free energy is $\Delta G = \Delta H - T\Delta S$ . What is the significance of the fact that the $\Delta H$ term has a positive sign, while the $\Delta S$ term has a negative sign?	Definitions of free energy, enthalpy, and entropy Translating parts of an equation into biological relationships
2. Precision	Rewrite the following sentence to add more detail, and explain why those details are important to the function of the system being described: "Evolution increases an individual's fitness by producing new mutations that will benefit that individual."	Distinction between the roles of mutation and natural selection Evolution as a population-level process Natural selection effected through current conditions, not potential future benefits
3. Connections within and between disciplines	Tyrosinase is an enzyme that catalyzes a key step in the production of the pigment melanin. Individuals who have two nonfunctional copies of the tyrosinase gene do not synthesize any tyrosinase and therefore produce no melanin. However, individuals with even one functional allele produce nearly as much melanin as those with two functional alleles. In what specific ways does this system's behavior resemble the SIR system you modeled? Explain.	Distinction between roles of enzyme and substrate Nonlinear behavior of processes that depend on interactions between separate components
4. Mature scientific judgment	Reflect on your overall understanding of water transport in vascular plants. Then choose the three specific components of this system that you believe would be <i>most</i> essential to modeling the overall process. Clearly explain your reasoning for choosing these three. Then explain your reasoning for <i>not</i> choosing at least one other important component of the system.	Cohesion of water molecules Role of leaves, stems, and roots in water transport Abstracting from a detailed to a more general understanding of a system Explaining functional connections between system components

Table 2: Questions for Assessing the Modeling Exercise's Learning Objectives

future lessons can reinforce model-based thinking without requiring the development of formal mathematical models. For example, students may explore cellular respiration by diagramming the flow of carbon atoms and electrons between key compartments (glucose, pyruvate, NAD<sup>+</sup> and NADH), then discussing the specific role of fermentation in the absence of oxygen. Similarly, genetics students might develop an informal model showing the stochastic and reversible binding of repressor protein to the *lac* operon, then discuss why low-level constitutive expression is essential to the induction of this operon. Conversations with colleagues both within and beyond biology can reveal many such applications of model-based thinking, enriching students' understanding of course material and illustrating the conceptual connections across the sciences with a minimum of additional time investment.

## Acknowledgements

The author thanks Pam Ryan, Kristin Jenkins, Yaffa Grossman, and two anonymous reviewers for their constructive suggestions and comments; David Hayes, Jacob Schrum, and Kyle Betts for demonstrating the modeling exercise's extensibility to an independent studies format; and the many undergraduate students and teaching assistants whose hard work and insightful questions both improved the exercise itself and helped identify several of the underlying lessons to be gained from modeling.

## References

- [1] J. Bransford, J. Franks, N. Vye, R. Sherwood (1989). *New Approaches to Instruction: Because Wisdom Can't be Told*. In S. Vosiadou & A. Ortony (Eds.), *Similarity and Analogical Reasoning* (pp. 470–497). New York: Cambridge University Press.
- [2] P. Burrowes. *A student-centered approach to teaching general biology that really works: Lord's constructivist model put to a test*. *Am. Biol. Teach.*, 65 (2003), No. 7, 491–502.
- [3] Centers for Disease Control (2004). *Smallpox Fact Sheet: Vaccine Overview*. Available online at <http://www.bt.cdc.gov/agent/smallpox/vaccination/facts.asp>; last accessed on 1/20/2011.
- [4] H. Ewing, K. Hogan, F. Keesing, H. Bugmann, A. Berkowitz, L. Gross, J. Oris, J. Wright (2003). "The role of modeling in undergraduate education". Pages 413427 in Canham CD, Cole JJ, Laurenroth WK, eds. *Models in Ecosystem Science*. Princeton (NJ): Princeton University Press.
- [5] S. Freeman. *Biological Science*, 4th edition. Pearson Benjamin Cummings, San Francisco, 2011.

- [6] R. Hake (1998). *Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses*. Am. J. Phys. 66 (1998), No. 1, 64–74.
- [7] J. Hodder, D. Ebert-May, and J. Batzli. *Coding to analyze students' critical thinking*. Front. Ecol. Environ. 4 (2006), No. 3, 162–163.
- [8] G. Johnson (2005). “Infectious Disease and Bioterrorism.” In PH Raven, GB Johnson, J Losos, and S Singer, *Biology, 7th edition* (Enhancement Chapter 33e). New York, NY: McGraw-Hill.
- [9] J. Jungck (2005). “Challenges, Connections, Complexities: Educating for Collaboration”. Pages 1–12 in *Math and Bio 2010: Linking Undergraduate Disciplines* (ed. LA Steen).
- [10] J. Jungck and J. Calley. *Strategic simulations and post-Socratic pedagogy: Constructing computer software to develop long-term inference through experimental inquiry*. Am. Biol. Teach. 47 (1985), No. 1, 11–15.
- [11] K. Kastens (2002). *My Top Ten Topics in Geoscience Education Research*, for a conference on “Bringing Research on Learning to the Geosciences.” [http://serc.carleton.edu/files/research\\_on\\_learning/KKtopten.pdf](http://serc.carleton.edu/files/research_on_learning/KKtopten.pdf) (accessed 1/17/2011).
- [12] W. Kermack and A. McKendrick. *A contribution to the mathematical theory of epidemics*. Proc. Roy. Soc. Lond. A 115 (1927), No. 772, 700–721.
- [13] E. Kitchen, J. Bell, S. Reeve, R. Sudweeks, W. Bradshaw. *Teaching cell biology in the large-enrollment classroom: Methods to promote analytical thinking and assessment of their effectiveness*. Cell Biol. Educ. 2 (2003), No. 3, 180–194.
- [14] J. Knight and W. Wood. *Teaching more by lecturing less*. Cell Biol. Educ. 4 (2005), No. 4, 298–310.
- [15] Lauenroth WK, Burke IC, and Berry JK (2003). “The Status of Dynamic Quantitative Modeling in Ecology”. Pages 32–48 in Canham CD, Cole JJ, Lauenroth WK, eds. *Models in Ecosystem Science*. Princeton (NJ): Princeton University Press.
- [16] M. Meltzer, I. Damon, J. LeDuc, J. Millar. *Modeling Potential Responses to Smallpox as a Bioterrorist Weapon*. Emerg. Infect. Dis. 7 (2001), No. 6, 959–969.
- [17] National Research Council. *Bio 2010: Transforming Undergraduate Education for Future Research Biologists*. National Academies Press, Washington D.C., 2003.
- [18] J. Reece, L. Urry, M. Cain, S. Wasserman, P. Minorsky, and R. Jackson. *Campbell Biology*, 9th edition. Pearson Benjamin Cummings, San Francisco, 2010.
- [19] SENCER: Science Education for New Civic Engagements and Responsibilities. ([www.sencer.net](http://www.sencer.net)). Accessed 13 January 2011.

- [20] J. Trempy, M. Skinner, W. Siebold. *Learning microbiology through cooperation: Designing cooperative learning activities that promote interdependence, interaction, and accountability*. *Microbiol. Educ.* 3 (2002), No. 1, 26–36.
- [21] Turner, MG (2003). “Modeling for Synthesis and Integration: Forests, People, and Riparian Coarse Woody Debris”. Pages 83–110 in Canham CD, Cole JJ, Laurenroth WK, eds. *Models in Ecosystem Science*. Princeton (NJ): Princeton University Press.
- [22] D. Udovic, D. Morris, A. Dickman, J. Postlethwait, P. Wetherwax. *Workshop Biology: Demonstrating the effectiveness of active learning in an introductory biology course*. *BioScience* 52 (2002), No. 3, 272–281.