From **Gal, I. & Garfield, J. B. (editors).** **The Assessment Challenge in Statistics Education**. IOS Press, 1997 (on behalf of the ISI). Pages 253-262. ISBN 90 5199 333 1. The References section is a separate chapter and is available on the Internet at http://www.stat.auckland.ac.nz/~iase/publications/assessbkref.

19. Probability Distributions, Assessment and Instructional Software: Lessons Learned from an Evaluation of Curricular Software

Steve Cohen
Richard A. Chechile

Purpose

While mathematics education guidelines have encouraged substantial change in the introductory probability and statistics curriculum, probability distributions still remain an important topic in a first course. In fact, just as software has made data analysis more accessible to students in introductory courses, it also offers new ways to teach probability distributions. However, these new teaching technologies, which emphasize active experimentation and interpretation of displays, also raise new questions. Just what do students see when they examine a display of a probability distribution? Do the displays really help students acquire a clear conceptual understanding? Can interactive exercises for related concepts like sampling distributions make good use of displays? Finally, can good assessment practices help us learn when displays are effective and when they might be confusing? This chapter will discuss some interactive, computer-based exercises that use and teach probability distributions, and consider how assessment can help address some of the important questions these new teaching technologies raise.

INTRODUCTION

The past decade has produced a great deal of thinking about the curricular goals for introductory courses in probability and statistics. Some of the thinking and research has been designed to demonstrate how deep the educational problems run. Garfield and Ahlgren (1988) have pointed out just how hard certain concepts are for students. Shaughnessy (1992), in a review of research on probability and statistics education, confirmed their findings. As a result of this research, calls for new instructional goals and methods of teaching have begun to redefine the introductory statistics curriculum. Cobb (1992) outlined several recommended changes. The recommendations included an emphasis on (a) statistical thinking rather than formulas and cookbook approaches, (b) more hands-on work with data, and (c) the use of interactive technologies to make data analysis more efficient and to help teach difficult concepts.

However, one traditional part of some introductory courses in probability and statistics has been a bit lost in this discussion. The role of the probability distribution in learning and teaching has been overlooked. In spite of the emphasis placed on hands-on data analysis and alternative methods for inference, probability distributions are, and likely will always be, a major part of a first course. Why? To begin, probability distributions, like the normal curve, are not found only in statistics courses and text books. They are part of the vocabulary for communicating basic ideas. Normal probability distributions are used to describe basic phenomena like grades and intelligence. While they might not be quite as ubiquitous as more common statistical concepts like averages and frequencies, they clearly extend beyond the classroom and consequently deserve a special place in the statistics curriculum.

Unlike data, probability distributions are formal theoretical models that describe the likelihood of a variable taking on a value (i.e., the binomial distribution) or a range of values (i.e., the normal distribution). They are useful for describing the behavior of sample means (i.e., a theoretical sampling distribution) and as such are important for teaching concepts (Moore, 1990) such as the central limit theorem and confidence intervals, as well as for doing basic statistics (NCTM, 1989). Their theoretical nature makes probability distributions a natural contrast to data which may help students develop a notion of stochasm. Understanding the *differences* between a data distribution and a probability distribution is one of the most profound insights a student can have. Indeed, the NCTM Curriculum and Evaluation Standards for School Mathematics (1989) encourages students in grades 5-8 to compare data with formal models to get a sense of stochasm.

Thus, even with a new, well placed emphasis on working with data, probability distributions still play an important part in teaching and learning basic statistics. Being able to interpret a probability distribution, and make well-reasoned claims about a variable by studying its probability distribution, is an important skill for almost any student in a first course in probability and statistics. How can we learn if we are successfully helping students achieve this goal?

INTERACTIVE EXERCISES INVOLVING PROBABILITY DISTRIBUTIONS

While all probability distributions have a mathematical form, only the most sophisticated students typically see, or choose to see, the associated equation. A more realistic goal, and one consistent with the kinds of changes called for by Cobb (1992), is to have students be able to interpret a display of a probability distribution and understand how it conveys probability.

One medium particularly well suited to helping students achieve these goals is statistical or curricular software. Software offers an excellent means to display probability distributions and to have students experiment with their different properties. Exercises may be focused on familiarizing students with properties of distributions themselves, or might use probability distributions to help illustrate other related concepts such as the central limit theorem. For instance, Figure 1 shows an interactive exercise from a program called ConStatS. The exercise is designed to help students understand the relationship between the parameters of a normal distribution and the shapes the distribution can take on. Students can change the parameters, and the resulting distribution is plotted on the same set of axes as the original distribution.

When displayed, probability distributions look deceptively simple. A set of probabilistic concepts is truly hidden in the plot of a simple function. It is easy to assume that students grasp the implied meaning. As a goal of an introductory course, it is important to know if a student can work with a probability distribution and see in them the rich and difficult concept of probability. This assessment goal is especially important if distributions are included as educational aids in other exercises (i.e., teaching sampling distributions.)

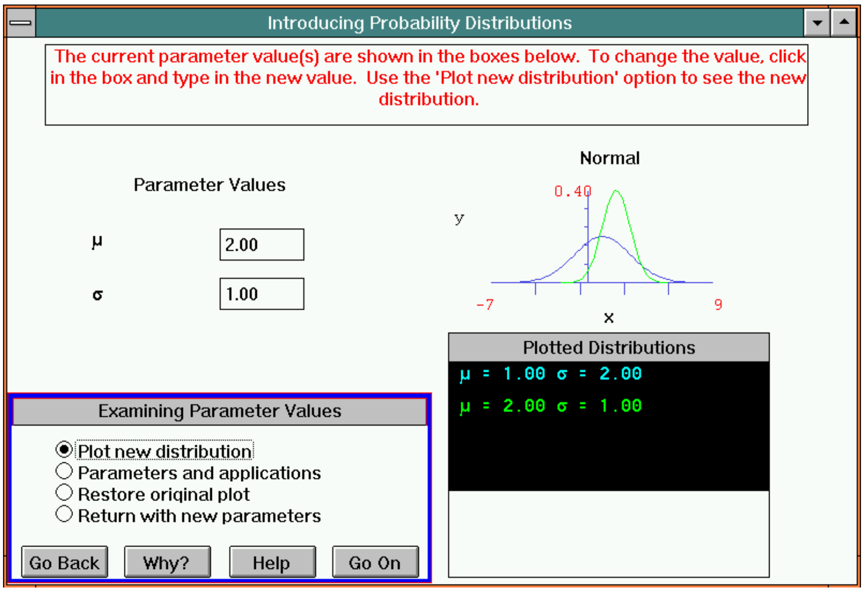


Figure 1: Exercise from ConStatS

ASSESSING STUDENTS' UNDERSTANDING OF PROBABILITY DISTRIBUTIONS

The remainder of this chapter will look at examples of instructional technology that teach or use probability distributions and consider the implications for assessing student understanding. In some instances the focus will be on learning probability distributions, and in other instances probability distributions will be used in examples designed to illustrate challenges in statistics assessment. Examples will be drawn from ConStatS (a set of interactive programs for helping students conceptualize topics covered in an introductory course (Cohen, Smith, Chechile, & Cook, 1994). Many of the insights in this chapter come from a three year, multi-site study of ConStatS (Cohen, Chechile, Smith, Tsai, & Burns, 1994). The scope of the evaluation (16 different introductory statistics courses at five different universities) should permit the results to generalize to most introductory courses and to most software for illustrating statistical concepts.

The chapter will proceed by first presenting some interactive educational exercises from ConStatS involving probability distributions. We will first look at the exercises and consider the educational goal(s). Then we will turn to the design and interpretation of assessment items, and look at how student performance on those items relates to overall conceptual understanding of material taught in a first course. The chapter will conclude with remarks about how assessing students' understanding of probability distributions connects to the bigger picture of assessment.

What is the software trying to teach?

ConStatS was created to help students gain a deep conceptual understanding of topics taught in introductory probability and statistics. This goal narrowed the assessment objectives in several important ways (Cohen et al., 1994). There is no emphasis in ConStatS on computational proficiency or remedial mathematical issues. For example, while the exercise shown in Figure 1 encouraged students to experiment with the parameters of various distributions and investigate the resulting forms, they were never required to calculate specific probabilities or become familiar with the mathematical form of the equation (though this is an option in the program). Consequently, almost all assessment items were designed exclusively to test conceptual understanding.

Yet even this focused educational goal permits a variety of questions that assess what students might have learned from experimenting with parameters of a distribution. For instance, Figure 2 shows a question used to assess a student's conceptual understanding of how the assigned mean and variance of a normal distribution affect its form. While this particular question tests conceptual understanding, other types of questions could assess similar kinds understanding. Specifically the question in Figure 2 is a *production* question. It requires a student to sketch a normal distribution with a given set of parameters. Alternatively, the question could have been designed to assess only *comprehension* by showing two distributions and asking the student to estimate the parameter values or to select from a set of possible options. It is important to look closely at the instructional exercise and consider what students are empowered to do by the software. The specific kinds of interactions can limit (or extend) learning, and the assessment items needs to reflect this. We will return to this issue later in the chapter.

Below is a normal probability distribution with mean = 2 and variance =1.
On the graph, sketch a normal distribution with mean = 1 and variance = 2.

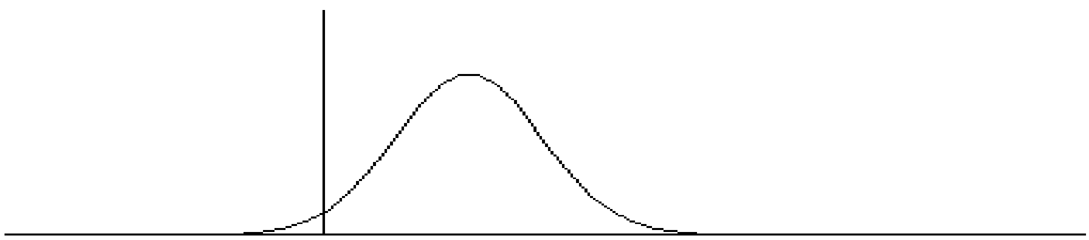


Figure 2: Item assessing understanding of the mean

HELP TEACH OTHER CONCEPTS

Some ConStatS programs use probability distributions to help teach other concepts. The two separate kinds of environments within ConStatS–programs treating probability distributions directly and those using them in context–lead to very different kinds of learning and require separate kinds of assessment items. For instance, there are several sampling programs that require students to use and interpret probability distributions as both population distributions and sampling distributions. Figure 3 shows a portion of the sampling distributions program with a normal probability distribution representing the population of human baby weights. In this environment, students are required to interpret a probability distribution, not construct one. Students who can effectively interpret a probability distribution in the context of a sampling exercise may not have a useful grasp of probability distributions in other contexts. It is possible, for instance, that a student could use information about a random variable (in this case human baby weights) to interpret the probability distribution in the sampling distributions setting. They may know that newborns are typically around seven pounds, and rarely weigh more than ten pounds and less than four pounds. This knowledge is consistent with a correct interpretation of the normal population distribution shown. However, the same student could have trouble interpreting a distribution without supporting information (i.e., for an abstract variable like X). This assessment issue emerges when considering the goal of the sampling distributions exercise. Students may be able to use intuitions about baby weights to make some sense of the population distribution, but then find themselves lost when looking at the sampling distribution of means about which they have no intuitions.

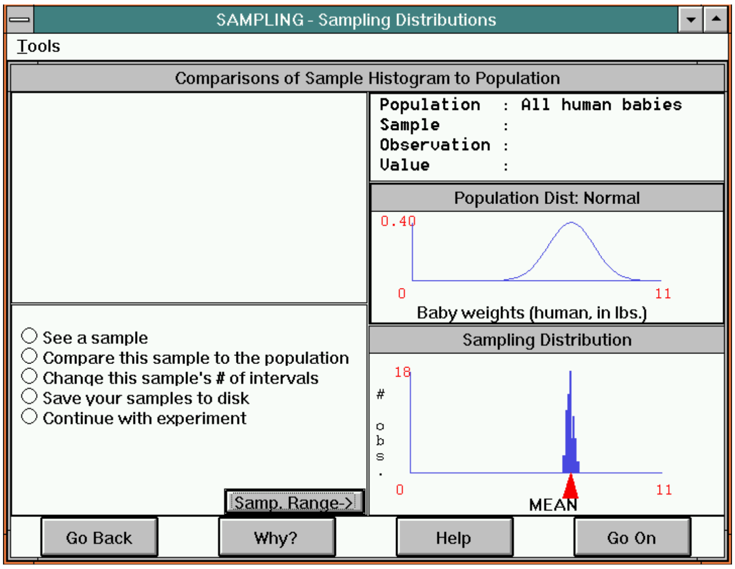


Figure 3: Sampling Distribution program

Exercises like the one in Figure 3 require students to interpret displays of probability distributions in order to learn statistical concepts like sampling distributions. Students who do not demonstrate a good understanding of spread when visually interpreting a normal distribution (i.e., do poorly on the question in Figure 2) may only partially grasp the concept of a sampling distribution. They may learn that sampling distributions describe distributions of statistics rather than individual data points, but may not come to understand the reduction in variance obtained when working with a sampling distribution rather than a population distribution.

The question in Figure 2 was included in the ConStatS evaluation. Students were typically able to shift the distribution correctly one unit to the left, but they had a great deal of trouble representing the adjusted variance (Cohen, Smith, Chechile, Burns & Tsai, 1996). The result illustrates the complex nature of using displays to convey concepts. Many students may have only a partial understanding of probability distributions. Ultimately, measuring a student's understanding of probability distributions is likely to require several well-formed questions that target specific properties and uses. Without such a detailed assessment, it may be impossible to learn why students have trouble with other key concepts like sampling distributions.

The effect of an interactive, visual medium

ConStatS, like most statistical software, makes extensive use of graphical representations. Many of the exercises in the software, like the one shown in Figure 1, require students to manipulate displays in order to learn and to interpret distributions. Even if students successfully learn how the parameters μ and σ influence normal probability distributions by doing experiments and studying resulting displays, the display may cause other confusions. For instance the normal distribution, as a theoretical construct, has a range from negative to positive infinity. The display, however, is limited. The range of the distribution may only be plotted to three or four standard deviations, and a student will only see the distribution(s) displayed over a finite range. The display tacitly conveys that the normal distribution is just a *finite*, smooth curve that represents some symmetrical data, rather than a distinct theoretical function that represents probability from negative to positive infinity. The limited display may reinforce this mistaken view. Figure 4 presents a way to assess student understanding about the range of the random variable. To supply a correct answer students must use a property of the normal probability distribution that is not evident from the display.

For each of the two probability distributions shown below, indicate the highest and lowest values the variable X can take on.

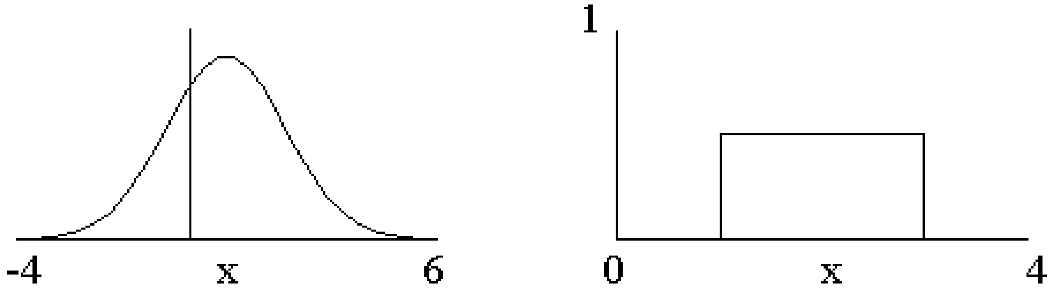


Figure 4: Item assessing the range of random variables

This assessment issue points to the potential limitations of using a visual and graphical medium to illustrate subtle ideas. Having students learn by investigating a representation of probability creates a special demand on assessment. It requires an assessment of the potential conceptual confusions that might emerge owing to the limits of a particular representation.

MATCHING ASSESSMENT ITEMS TO INSTRUCTIONAL MATERIALS

If assessments of new instructional methods like ConStatS are going to provide feedback on how the methods should be improved, the assessment item needs to match instructional method very closely. Earlier in the chapter the discussion centered on a question for assessing a student's understanding of the parameters for the normal distribution. Two different kinds of questions were considered: a production question and a comprehension question. It is difficult to determine which of these two kinds of questions is the best match for the exercise in the software. Students using the exercise in the software (Figure 1) first selected an existing distribution and then saw the values of the parameters. They were then asked to modify the parameters, and review the resulting distribution. In this exercise, students are (in some sense) producing a normal distribution by defining a new set of parameters and having the software draw the new form of the distribution. Owing partially to a limitation in the instructional method, students are never actually required to sketch or otherwise produce a distribution by hand (as required by the question). Consequently, some students may not confront the kinds of subtle decisions that sketching a graph demands. Having students sketch the graph may force them to pay much closer attention to inflection points, the total area covered, the range of the plot, and other more detailed properties of the distribution and the display. Sketching is a very different educational exercise than creating displays by computer, yielding a different set of skills and appreciation for the distribution. It is the kind of educational exercise that is often omitted when students move from pencil and paper exercises to software.

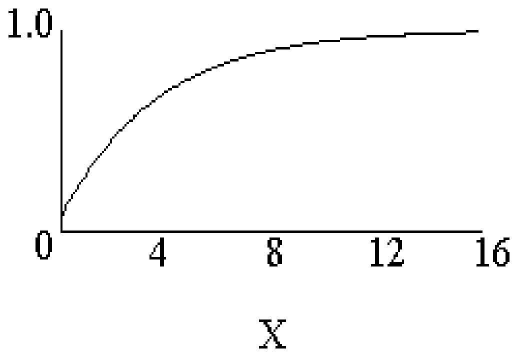
How does reconsideration of the exercise in Figure 1 influence interpretations of student responses to the question in Figure 2? In this instance, a production question might not match the instructional method quite as well as a comprehension question. Students who do not do well on the question in Figure 2 may have a gained a basic understanding of parameters and forms of the normal distribution from the exercise, enough to answer a comprehension question correctly. Yet their understanding may not generalize well to slightly more demanding tasks.

INTERPRETING OPEN-ENDED QUESTIONS

Production questions, like the one in Figure 2, raise the additional issue of the value of open-ended questions. When using a new instructional medium like software to help teach sophisticated concepts, open-ended questions can be very valuable. Students come to a question with a wide range of skills and knowledge and offer responses that are difficult to anticipate. Often these unanticipated responses offer clues to conceptual confusions. An example from the

ConStatS evaluation should illustrate the point. The question in Figure 5 required students to construct a uniform probability density function that matched the cumulative density function. Several students drew "mostly" uniform distributions with curved tail areas from a normal distribution at the ends. It appears that at least some students have trouble abandoning a normal model when considering other probability distributions. Students who demonstrate a good conceptual understanding of normal distributions may display subtle misunderstandings when asked to interpret other probability distributions. Questions with these kinds of unanticipated answers may be very difficult to score, but they provide evidence about what kinds of misunderstandings plague students. When many students exhibit the same kind of misconception, it may indicate a strong cognitive bias that is difficult to address by instruction (Smith, diSessa, & Roschelle, 1992; Cohen, et al., 1996). Production questions allow students to demonstrate common misconceptions through patterns of errors that are often difficult to anticipate.

Based on the cumulative density function below, what range of values is the variable X most likely to take on?



1) 0   <= X < 4
2) 4   <= X < 8
3) 8   <= X < 12
4) 12 <= X < 16

Figure 5: Probability density function

In many instances, it is possible to predict the kinds of mistakes students might make if they lack a certain understanding or have a certain misunderstanding. Under these circumstances, short answer questions that play on the misconceptions are very useful for separating students who have a good understanding from those who do not. Figure 6 shows a cumulative density function and requires students to indicate which range of values the variable is most likely to take on.

It is possible students might misinterpret the plot as a probability density function and select the option corresponding to the largest area under the curve. Another possibility is that students believe that the height of the ordinate, which represents probability, indicates the likelihood of any given value being observed (i.e., they read it as if it was a bivariate plot). Students with either of these misconceptions will almost certainly select the last option, 12<=x<=16.

Unfortunately, the question does not help distinguish which of the anticipated misconceptions is serving as the chief impediment to successful interpretation of the graph. Items like the one in Figure 6 must be improved to provide more precise feedback about the particular misconceptions operating.

The questions in Figures 5 and 6 also illustrate another principle for evaluating conceptual understanding, transference. The key is to instantiate the concept under evaluation in a scenario unfamiliar to the student. Many students may be familiar with normal distributions, and with the relationship between the normal probability density function and its cumulative density function. They may have seen the relationship in a book, or even encountered it in another class. Their understanding may be limited to a rote visual recollection rather than a true understanding of the relationship. Using an unfamiliar distribution helps reduce the chance that a student is relying only on rote memory, rather than reasoning, to answer the question.

Construct a probability distribution which corresponds to the cumulative probability distribution shown below.

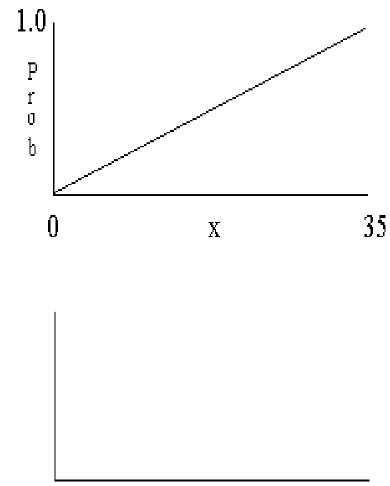


Figure 6: Cumulative density function

IMPLICATIONS

As the goals of teaching probability and statistics move from calculating numbers to reasoning and interpretation, and the teaching medium changes from blackboards to software, assessment practices must also adapt. The kind of assessment issues raised are not unique to ConStatS. Many computer programs used in the statistics curriculum, whether designed for doing statistics (like SPSS or Data Desk) or teaching statistics, use representations in similar ways. For instance, Rubin and Rosebery (1990) describe an interactive tool for the pre-college sector called ELASTIC, one part of which invites students to interactively stretch bars on histograms and examine changes in statistics. As in ConStatS, the pedagogy involves interacting with representations and drawing inferences about related statistical concepts. This pedagogy, and its educational goals, requires assessment practices that examine the components of the pedagogy and help us understand the effectiveness of representations and interactions. What questions should drive the construction of appropriate assessment strategies?

When considering exercises like the ones used in ConStatS, several issues emerge. When students are shown a probability distribution in an instructional exercise, what exactly do they see? What concepts does it evoke? And what concepts must it evoke for students to use it in a learning exercise? Assessment items that provide reliable answers to these questions need to be crafted with a great deal of care. For instance, in the sampling distributions program pictured in Figure 3, the hope is that students, when seeing the population distribution, use concepts of probability to interpret sampling processes. The representation of a probability distribution needs to be a window into probabilistic concepts. While the set of concepts in a student's mind may not include all the formal properties of probability and probability distributions, they should permit the student to realize why a probabilistic definition of the variable is salient and how using the variable in a population setting is different from using it when it is described by sample data.

Do these visual, non-mathematical representations open the necessary conceptual window? At least two questions in the ConStatS assessment cast some doubt on whether displays of probability distributions always fulfill this role. In the question shown in Figure 4, students are asked to identify the highest and lowest values the variable described by the distribution could take on. Many students selected the range from the display, rather than true limits of the distribution. On another question, students were shown a normal population distribution that represents the weight of newborn cats and asked what the disadvantages of using a normal distribution might be to describe feline birth weight. Many students claimed that the normal distribution would *not account for outliers* (Cohen, et al., 1996). It appears that the limits of the display were taken to define the limits of observable data. This interpretation would be appropriate if students believed that they were looking at data. In this instance, the pattern of errors across questions indicates that some students may be confusing displays of data with displays of probability distributions. Why might this be happening?

There are several possible interpretations. Some students may not have well-formed ideas about probability, and interpret probability distributions as univariate data. It may also be that students do have separate, well-formed representations of probability and data, but probability distributions do not invoke the correct set of concepts. Finally, it may be that in spite of these errors, students do bring a suitable subset of probabilistic concepts to the instructional exercise. If either of the first two options is true, especially the second, then using probability distributions (or any abstract graphical representation) in an instructional exercise may be a delicate matter.

A principal goal of any assessment must be to help identify this kind of subtle confusion. If these subtle confusions can be identified and corrected, displays of probability distributions can be used to successfully illustrate a number of important concepts. The cumulative wisdom gained from good assessment will help teachers know when they can draw a probability distribution and have their students understand its meaning.

One final point to emphasize here is that results from two different kinds of questions supported each other in illustrating a misunderstanding. The question on high and low values requires only that students interpret a plot, while the question on population distributions for newborn kittens has a generative component (i.e., students must produce an explanation.) Assessment items and practices can and should complement each other. Different kinds of assessment items will help provide a more complete profile of what students are and are not learning, and why. There is no one correct venue for assessment. Student explanations, like those required for the question on population distributions, can come in informal settings, like computer labs, or as part of formal assessment items. Even short-answer multiple-choice questions, like the one in Figure 6, can probe conceptual understanding. What is important is to create a variety of measures that complement each other, and which offer opportunities to examine student understanding and improve curricula.