Susan A. Peters

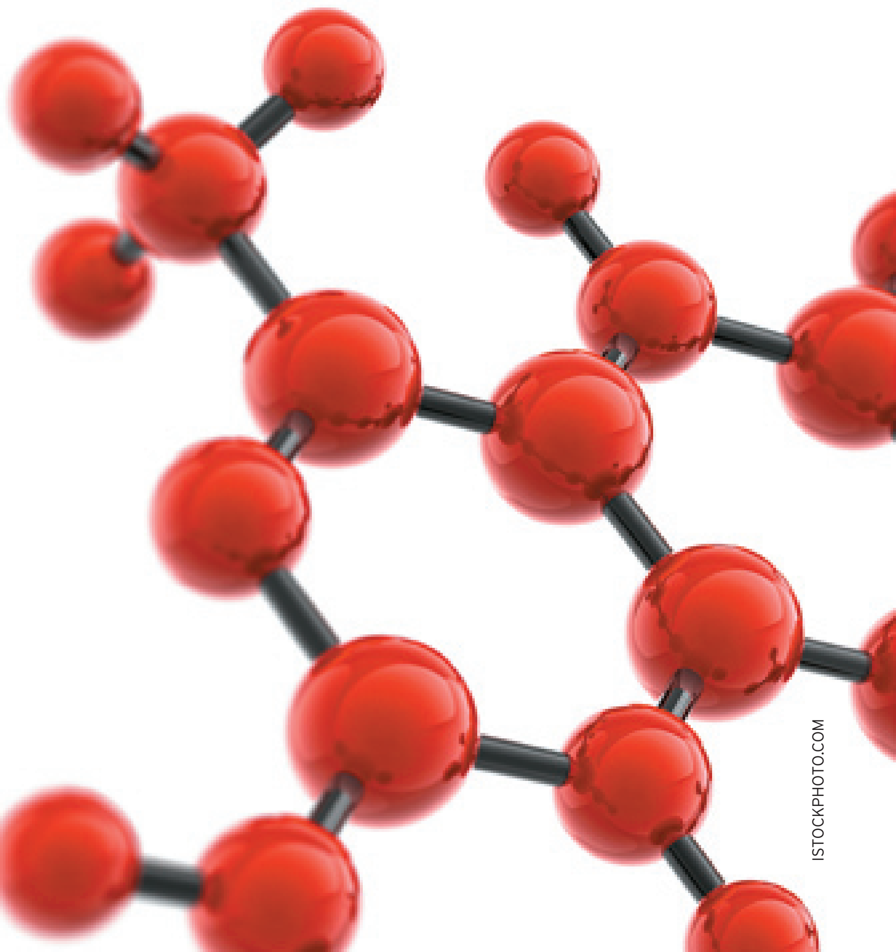
Engaging with the Art &

Statistics uses scientific tools but also requires the art of flexible and creative reasoning.

The trouble with statistics is that it is not mathematics.
—David S. Moore (1997, p. 93)

How can statistics clearly be mathematical and yet distinct from mathematics? The answer lies in the reality that statistics is both an art and a science, and both aspects are important for teaching and learning statistics. Here we describe artistic and scientific facets of statistics and use a task to illustrate how teachers can encourage students to incorporate elements of both in their reasoning.

&Science of Statistics



ISTOCKPHOTO.COM

HOW IS STATISTICS BOTH ART AND SCIENCE?
Statistics is a mathematical science in that it applies mathematical theories and techniques. Mathematics provides the theoretical means for justifying statistical methods as well as the procedural tools for implementing them. Methods for collecting, analyzing, and interpreting data combine to form the science of statistics. For example, we follow procedures to produce graphical representations of data and calculate summary statistical measures. Graphing and calculating involve scientific, statistical methods formulated from mathematical techniques.

However, we need more than science to decide which statistical measures best describe data and which statistical displays best reveal defining characteristics of data. Insightful, subjective decisions made during the processes of collecting, analyzing, and interpreting data to form conclusions within a given context constitute the art of statistics.

Consider a set of data that contains the heights of male and female adults. Men generally tend to be taller than women, a fact that causes a distribution of adult heights to be bimodal. A histogram may or may not reveal the bimodal nature of the data. The histogram in **figure 1a** displays the heights of 50 adult males and 50 adult females, and the histogram in **figure 1b** displays the heights for the same 100 adults. The histogram in **figure 1a** reveals a bimodal distribution, but the histogram in **figure 1b** does not.

Choosing bin widths and intervals that reveal important data characteristics in a histogram requires an artful touch. In general, histograms are constructed with between five and fifteen bins. If we use large bin widths or if we arbitrarily choose intervals, we may fail to reveal distinctive characteristics of data. Technology tools allow us to create multiple histograms using a variety of bin widths and intervals to aid us artistically in selecting the "best" picture for the data. Repeated explorations with data provide opportunities for students to develop the flexible and creative reasoning that typifies the art of statistics.

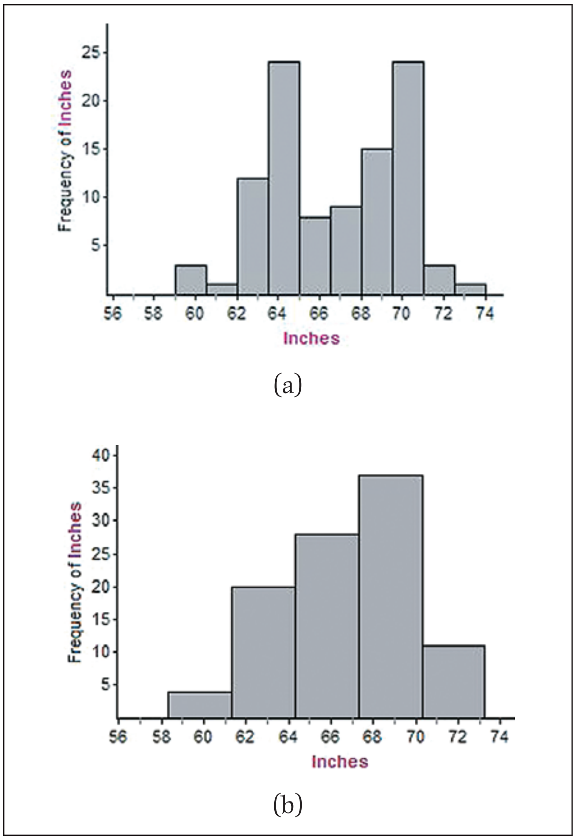


Fig. 1 The bimodal nature of the data displayed (**a**) can be hidden by changing the bin widths (**b**).

STATISTICS AS A PROBLEM-SOLVING PROCESS

Although statisticians use the art and science of statistics to collect and analyze data, statistics instruction traditionally has focused on the science of statistics (Groth 2006). Students calculate values for statistical measures or construct graphical data displays, but they spend little time discussing the appropriateness of using particular summary measures or discussing the meaning of the measures (Friel, O'Connor, and Mamer 2006). When making subjective decisions for displaying data, students rarely consider what graphical displays reveal or conceal about data. Today's students are expected not only to possess technical skills but also to interpret graphs, reason from measures, use sound data collection methods, and infer characteristics of populations from samples, all with an artistic eye on the context of the data (Moore 1997). Attention currently is focused on finding ways to establish this new tradition in statistics instruction.

One tool for considering how the art and science of statistics might appear in pre-K–12 mathematics education is the Guidelines for Assessment and Instruction in Statistics Education (GAISE) framework (Franklin et al. 2007). The GAISE report presents a coherent and focused vision for statistics curricula that centers on a four-step statistical problem-solving process and the role of variation within that process. The artful nature of statistics can be found in contextual considerations of variation, which complement the science of statistics embodied within the problem-solving process.

The first stage of the statistical problem-solving process involves formulating a statistical question and anticipating possible sources of variation that might interfere with answering the question. In the second stage, data are collected through a method that allows the question to be answered while reducing the effects of anticipated sources of variation. Data analysis, the third stage, includes quantifying the variability in data with measures such as the standard deviation. Analysis also includes accounting for variability when selecting a model appropriate to the data, such as a normal distribution. The fourth and final stage of the problem-solving process includes allowing for variability when interpreting results in context and in response to the statistical question. In addition to presenting a framework for statistics curricula, the GAISE report (Franklin et al. 2007) presents a useful framework for designing effective statistics lessons. As an example of how we can use the statistical problem-solving process in teaching statistics at the secondary level, we will use a task called the Consultant task.

THE PROBLEM-SOLVING PROCESS IN ACTION

The Consultant task (see **fig. 2**) was designed to elicit statistical reasoning in each of the four problem-solving stages by purposefully omitting statistical information needed to reason about the problem posed in the task description. The task will illustrate how the scientific and artistic facets of statistics are intermingled within the problem-solving process. The task has been presented in a professional development activity with secondary school statistics teachers, and some of their reactions are included in the discussion to underscore how important it is to keep one eye on the art and one eye on the science. Students too can use the task as a course-culminating activity or as an activity to be revisited throughout the year as they encounter concepts and methods that can help them reason with greater sophistication. The raw data for the task appear on page 502.

Formulating Statistical Questions

To a large extent, formulating statistical questions is an artistic endeavor. Students need to sift through multiple factors that may interfere with their ability to pose a question that can be answered using statistical methods. The Consultant task contains an implicit statistical question, but the language in the description presents an opportunity for students to wrestle with extracting the question of whether a difference in scoring exists. For example, on reading the task statement, some teachers expressed confusion regarding the administrators' stated desire to improve students' test scores; they were troubled by the extraneous information presented in the task description. Acknowledging the need to sort out the extra information, we can focus students' attention on the administrators' question by asking what question the administrators want answered or what question can be answered from the information given. For students to be able to engage in the problem-solving process independently, we need to give them opportunities to consider issues related to the artistic act of formulating statistical questions.

Collecting Data

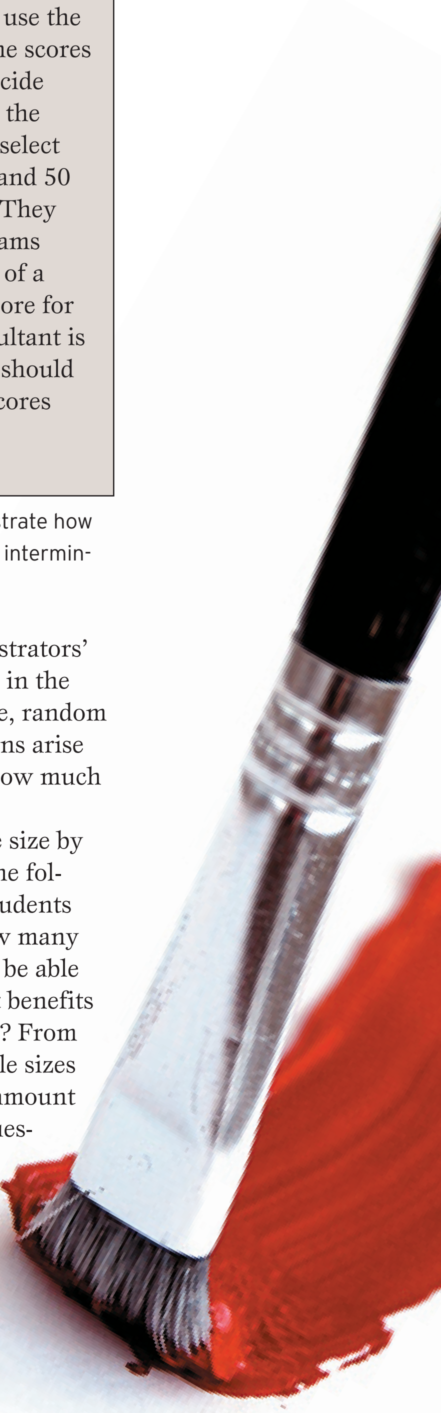
For the Consultant task, the problem-solving stage of data collection was completed by the administrators. The task provides little information about how the administrators selected the exams and thus presents a venue for students to question the selection process and discuss how the study could

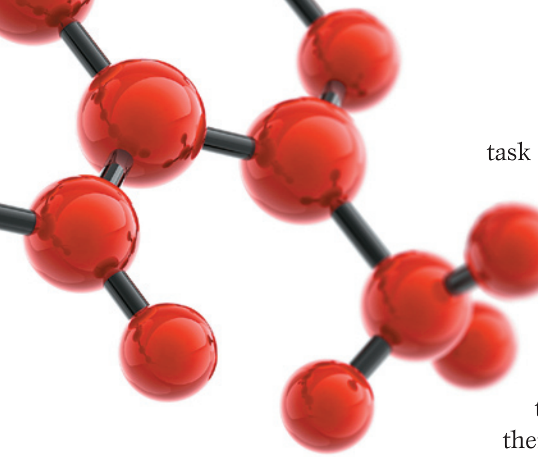
> To improve students' test scores on state assessments, administrators from a large school district require students to take practice exams. Two outside consultants create and score the open-ended questions from these exams. Although both consultants use the same rubric to score student responses, the administrators suspect that the consultants do not interpret and apply the rubric in the same way, resulting in differences in scores between the exams scored by the two consultants. The consultants' contracts with the district are up for renewal, and the administrators are trying to decide if they should be renewed. They decide to use the most recent practice exam to compare the scores assigned from each consultant and to decide whether there is a difference in the way the exams were scored. The administrators select 50 exams scored by the first consultant and 50 exams scored by the second consultant. They find that the average score for the 50 exams scored by the first consultant is 9.7 (out of a possible 15 points), while the average score for the 50 exams scored by the second consultant is 10.3 (out of a possible 15 points). What should the administrators conclude about the scores assigned by these two consultants?
> Source: Peters (2009)

Fig. 2 Solutions to the Consultant task will illustrate how the scientific and artistic facets of statistics are intermingled within the problem-solving process.

have been designed to address the administrators' question better. Important considerations in the data collection process include sample size, random selection, and control. These considerations arise in response to questions about how and how much data should be collected.

We can help students deal with sample size by having them consider questions such as the following, which are appropriate whether students have studied formal inference or not: How many exams should the administrators select to be able to answer their statistical question? What benefits could result from selecting a small sample? From selecting a large sample? In general, sample sizes are determined by finding the minimum amount of data needed to answer the statistical question with some degree of certainty. Scientific methods exist to provide some guidance, but finding a balance between collecting too little data to draw conclusions and too much data to be feasible presents an artistic challenge. Many teachers who worked through the Consultant





task noted that the sample size of 50 was sufficient to use common statistical methods of comparison (namely, the central limit theorem and traditional inferential methods).

If students do not mention the need for random selection, then the issue should be raised by asking questions about how the exams could be selected to be able to answer the statistical question and by presenting sampling strategies for students to critique. We choose sampling methods to achieve the highest probability of obtaining a representative sample within the practical constraints of the setting.

Random sampling is one method used to achieve this high probability, even though it cannot eliminate the possibility of selecting a nonrepresentative sample. For example, if 50 exams are selected from those graded in the first hour, when a consultant may be reading each exam carefully, or if 50 exams are selected from those scored at the end of a long and exhausting day, when a consultant may become careless, the exams may not be representative of all exams scored by the consultant. Random sampling ensures a low probability for selecting either of these two extreme sets of exam scores or other extreme samples from among the many possible sets of 50 exam scores.

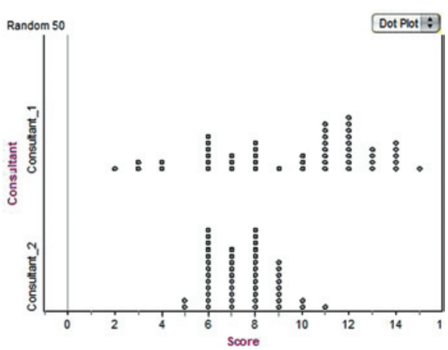
When discussing data collection methods, teachers raised a concern about whether the scores were randomly selected, although few considered how the exams were allocated to the consultants before scoring. Random sampling of scored exams is not sufficient for comparing consultants' scores. It may not reduce the chance for bias that may exist if, for example, one consultant scores the exams of students from one teacher while the other consultant scores the exams of students from a different teacher. If we fail to pay attention to how the exams were selected for the consultants to score, we may find a difference in scores that is attributable to teachers rather than to scoring. To reduce the chance for bias, we should randomly assign exams to each consultant for scoring. Random assignment of exams to consultants and random selection of scored exams should allow us to notice potential differences in the way the consultants graded the exams. As teachers, we can raise the issue of random selection with students by presenting scenarios such as the Consultant task and asking students to reason about the appropriateness of the data collection method used. Other design methods such as matched pairs may work even better.

Students should consider other data collection methods and how administrators might benefit from these methods. Most teachers who were asked how they would design the Consultant task suggested using a matched-pairs design by having both consultants score the same exams. Using scores from the same exams shifts attention from the scores as sets of numbers to differences in scoring between the consultants. Any variation we see could then be attributed to the consultants rather than to the students. Because the administrators want to focus on how the consultants use the rubric, they can control the effect of variation in students to reveal a clearer picture of any differences in consultants' scoring.

Determining appropriate sample sizes, introducing random selection into data collection, and finding ways to control variation are examples of artistic decisions associated with data collection. The art comes from selecting the design that best allows us to answer the statistical question; the science occurs in implementing the process. When students critique designs such as the study design described in the Consultant task, they learn to recognize the complexity of considerations they encounter when designing a study.

	Consultant 1	**Consultant 2**
Mean	9.7	10.3
Standard Deviation	3.376	20.205

(a)



(b)

Fig. 3 Additional information about the variation in scores, either summary statistics (**a**) or dot plots (**b**), was needed to move forward with the task.

Analyzing Data

Students seek answers for the statistical question during the data analysis stage, the most scientific stage of the problem-solving process and the one with which students are most familiar. The Consultant task description does not contain enough information to answer the administrators' question statistically; students need to decide what additional information is needed. Initially, the only available information is the average of 50 exam scores for each consultant. Teachers who engaged in the task quickly recognized that they needed to know something about the variation in scores to determine whether the difference was larger than could be expected by chance. They asked for either the summary measures (see **fig. 3a**) or the data (as shown in the dot plots in **fig. 3b**).

An interesting discussion about artistic and scientific elements of analysis can result from examining the standard deviations of the consultants' exam scores. Students who pay attention strictly to the science of statistics and who have experience with inferential techniques may be tempted to launch into a significance test, a strategy that also tempted the teachers who saw the summary measures but not the graphs. A significance test led participants to the conclusion that there is not enough evidence to show a significant difference in the consultants' average scores.

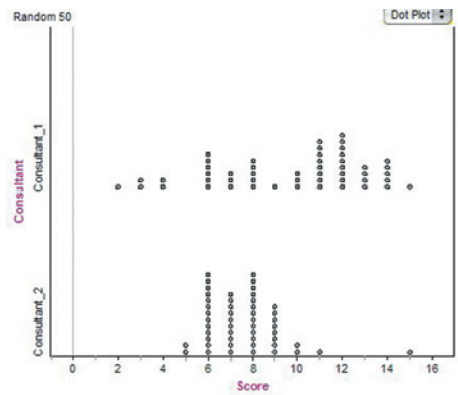
This conclusion, however, does not consider the context of the data. If we think about standard deviation as a measure of the approximate average absolute deviation of data from the mean, then the standard deviation value of 20.2 for consultant 2's scores is not possible. The average absolute deviation in scores from a mean of 10.3 cannot be 20.2 if the exams are scored on a scale from 0 to 15. Part of the art of statistics is recognizing the limitations of reasoning exclusively from scientifically calculated summary measures and considering restrictions that may exist in particular contexts.

Many teachers who reached this impasse as a result of examining the summary measures asked to see the data. After examining the dot plots, they suggested that the value of the standard deviation for consultant 2's scores was less than the value of the standard deviation for consultant 1's scores. Consultant 1's scores certainly seem to be more varied than consultant 2's. In fact, the task was designed to create this exact dilemma through a "data entry error" for one of the scores assigned by consultant 2. A value of 150 was entered for an exam score of 15. The value was not displayed on the dot plot because the dot plot was limited to scores from 0 to 15.

This data entry error presents an opportunity for students to reason about the two consultants' scores by comparing summary values and dot plots. Only by considering the context of the data artistically and by making scientific comparisons among representations will students realize that a problem exists. This error also presents an opportunity for

	Consultant 1	Consultant 2
Mean	9.7	7.6
Standard Deviation	3.376	1.726

(a)



(b)

Fig. 4 Once the data entry error has been corrected, summary statistics (**a**) and dot plots (**b**) change.

Raw Data for the Consultant Task			
Consultant 1		**Consultant 2**	
Score	Frequency	Score	Frequency
0	0	0	0
1	0	1	0
2	1	2	0
3	2	3	0
4	2	4	0
5	0	5	2
6	6	6	13
7	3	7	10
8	5	8	13
9	1	9	8
10	3	10	2
11	8	11	1
12	9	12	0
13	4	13	0
14	5	14	0
15	1	15	0 (1 in full set)

students to consider how the outlier affects the values for both the mean and the standard deviation.

Examining the summary values (see **fig. 4a**) and the corrected displays (see **fig. 4b**) reveals two distributions with different centers (means) and different variations (standard deviations). Many teachers suggested conducting a test to determine whether a significant difference existed between the means, but few suggested conducting some type of significance test for the standard deviations—perhaps a sign of how particular methods dominate what most of us have experienced as learners and teachers. Even without significance tests, the differences in summary measures provide an opportunity to discuss what the administrators meant by the term *different* or how the summary measures and graphs can be used to help students think about whether there is a difference in the averages and standard deviations for the scores assigned by the two consultants.

Interpreting Results

In addition to giving students a chance to use data analysis tools, the Consultant task context allows them to engage in discussion about the administrators' real issue: How much variation in scores between the two consultants is too much variation? The consultants' scores can be compared in a variety of ways, including comparisons using summary measures, graphs, or significance tests. For example, the standard deviation of the scores assigned by consultant 1 is almost twice that of consultant 2. The distribution of scores for consultant 2 is fairly symmetric and mound shaped and contains only one score outside the interval from 5 to 11, whereas the distribution of scores for consultant 1 is left skewed and contains scores in the inverval from 2 to 15. A two-sided, two-sample t test yields a p-value of .0002. No matter which of these types of comparison is made, students are likely to suggest that the consultants do not use the rubric in the same way—the differences seem to be greater than we would expect from random samples of exams scored in the same manner.

Whether we use a significance level or substantial differences in summary measures or other distributional characteristics to decide at what point a difference exists beyond what we would expect, the precise criteria we use suggest the presence of artistry. Artistry can be found in using the context and considering the repercussions for an erroneous conclusion. In contrast, implementing the criteria is a largely scientific endeavor. Interpreting results and drawing conclusions combines the scientific interpretation of the criteria along with the artistic consideration of the context.

CONCLUSION

The open-ended Consultant task paired with focused questions allows students to engage with both the art and the science of statistics. The task is ideal for combining artistic observatations and scientific calculations within the four stages of the statistical problem-solving process in a way that aligns with the vision of statistics education presented in the GAISE report (2007). Using this and similar tasks can advance a new statistics education tradition in secondary school mathematics education—a tradition that blends art with science.

ACKNOWLEDGMENT

This article is based in part on the author's doctoral dissertation, which is supported in part by the National Science Foundation under Grant ESI-0426253 to Pennsylvania State University. The ideas expressed here are those of the author and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

Franklin, Christine, Gary Kader, Denise Mewborn, Jerry Moreno, Roxy Peck, Mike Perry, and Richard Scheaffer. "Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report: A Pre-K–12 Curriculum Framework." Alexandria, VA: American Statistical Association, 2007.

Friel, Susan N., William O'Connor, and James D. Mamer. "More Than 'Meanmedianmode' and a Bar Graph: What's Needed to Have a Statistical Conversation?" In *Thinking and Reasoning with Data and Chance*, 2006 Yearbook of the National Council of Teachers of Mathematics (NCTM), edited by Gail F. Burrill and Portia C. Elliott. Reston, VA: NCTM, 2006.

Groth, Randall E. "Engaging Students in Authentic Data Analysis." In *Thinking and Reasoning with Data and Chance*, 2006 Yearbook of the National Council of Teachers of Mathematics (NCTM), edited by Gail F. Burrill and Portia C. Elliott. Reston, VA: NCTM, 2006.

Moore, David S. "Probability and Statistics in the Core Curriculum." In *Confronting the Core Curriculum*, edited by John Dossey. Washington, DC: Mathematical Association of America, 1997.

Peters, Susan A. "Developing an Understanding of Variation: AP Statistics Teachers' Perceptions and Recollections of Critical Moments." PhD diss., Pennsylvania State University, 2009.

SUSAN A. PETERS, s.peters@louisville.edu, is an assistant professor in the Department of Teaching and Learning at the University of Louisville. She teaches prospective middle school and secondary school mathematics teachers and is interested in statistics education and mathematics teacher education. SUSAN A. PETERS