Michael Todd Edwards

# Who Was the Real William Shakespeare?

*Connecting language arts and mathematics, students use data analysis and readability measures to identify the Bard.*

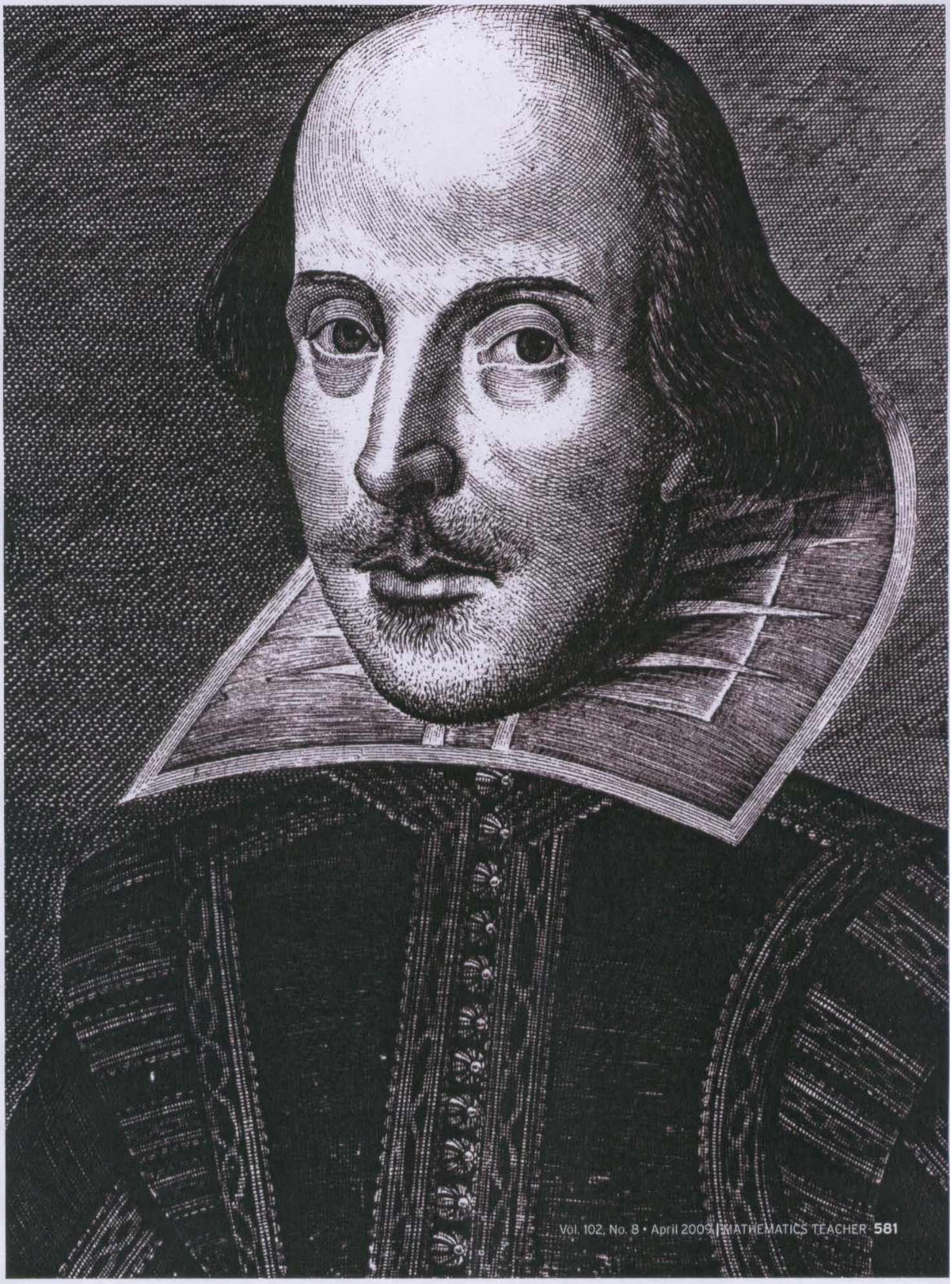
In recent years, calls to incorporate reading and literature more meaningfully into secondary school mathematics instruction have become pronounced (Borasi and Siegel 2000; Martinez and Martinez 2001; Mower 2003). Historically, however, literature and mathematics have been connected less frequently at the high school level than in the early or middle grades. Making connections between literature and mathematics in high school courses can serve several crucial purposes: addressing the weak reading skills of today's students, using technological tools to foster interdisciplinary study in schools, and capitalizing on synergistic learning outcomes that are often realized from interdisciplinary study.

This article presents an interdisciplinary project I have used with students to connect reading and mathematics instruction at the secondary school level. Using a data analysis approach, students examine Shakespearean sonnets in a course entitled Functions, Statistics, and Trigonometry while simultaneously studying the same works from literary and historical points of view in an English literature course. As they learn the basics of testing statistical hypotheses, the students use statistics to determine the likelihood that William Shakespeare was actually the pen name of Edward de Vere, a long-standing hypothesis popular among linguists and historians and promulgated by recent works such as *The Shakespeare Enigma* (Dawkins 2004).

## IDENTIFYING WILLIAM SHAKESPEARE

As ninth and tenth graders study Shakespeare in language arts classes, they are given the article "Hunting for Good Will: Will the Real Shakespeare Please Stand Up?" (Satchell 2000), which discusses an ongoing controversy surrounding Shakespeare's identity. The author provides evidence supporting the view that William Shakespeare was actually a pseudonym of Edward de Vere (1550–1604), the seventeenth Earl of Oxford, a recognized poet and playwright of the time. Although the debate about Shakespeare's identity has continued for centuries, until recently statistical analysis—including comparisons of de Vere's poems with those of Shakespeare—were simply beyond the grasp of most scholars. However, with the increased popularity of graphing calculators, statistical investigations of Shakespeare's writing now lie within the reach of typical high school students.

Technology allows students a chance to engage in authentic research, here comparing literary works of various authors in a manner similar to Johnson (1994) and others. Specifically, students collect and analyze data collaboratively in an attempt to answer contemporary research questions regarding Shakespeare's identity.



## IMPLEMENTING THE LESSON

Along with the Satchell (2000) article, students are given the article "Comparing Texts and Identifying Authors" (Johnson 1994) and eleven poems to analyze—ten by Shakespeare and one by de Vere. As a homework exercise, students construct for each assigned poem a data table that includes the number of letters (word length) and the number of syllables for each word in the poem. This task may be accomplished either manually or by using technology.

Although the initial data collection activity may be accomplished without a computer, my students typically use a Microsoft Word® macro to automate these tasks. To do this, students open a Microsoft Word .dot file containing the macro (Windows and Mac OSX versions of the macro are available for download at www.users.muohio.edu/edwardm2/macros) and then type or paste the text they wish to analyze. After highlighting the desired text within Word, students click a macro button that appears in the application toolbar. These steps are suggested in the screen shot of the open Mac OSX .dot file shown in **figure 1**. The macro highlights words from the selected text one at a time. After students enter the number of syllables of the selected word, the macro advances to the next word. A screen shot of the running macro is shown in **figure 2**.

Once the macro processes all the words, Flesch-Kincade grade-level statistics for the passage are calculated, displayed to the Calculate Readability window, and copied to the Word clipboard. Data may then be pasted directly into Microsoft Excel, Minitab, Fathom—or any other comparable data analysis application—for further analysis.

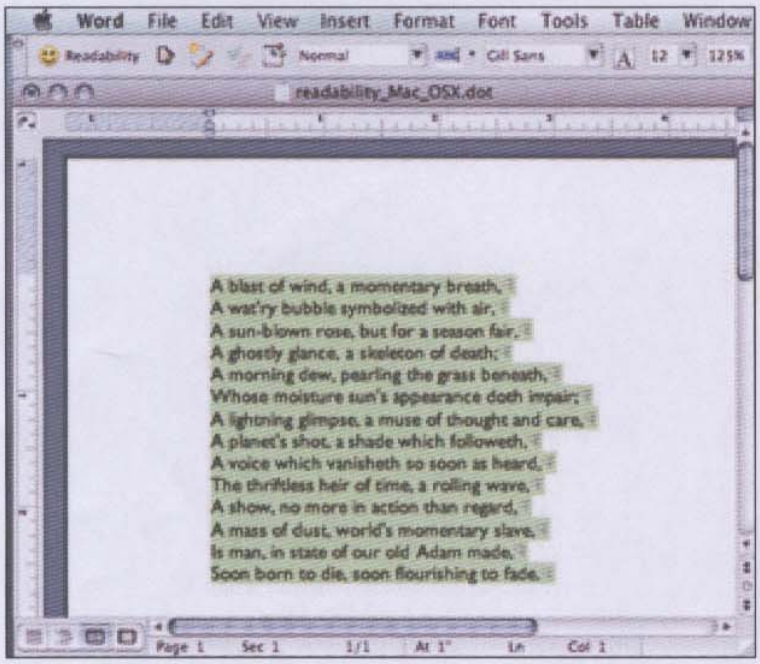



**Fig. 1** A sample sonnet highlighted within the Microsoft Word .dot file. After highlighting the text, students click the Readability button in the toolbar to execute the macro.

### *Initial Analysis of Aggregated Data*

To obtain data relating to each of Shakespeare's 154 sonnets and each of de Vere's 25 known poems (Shakespeare Fellowship 2006), I assign different sets of poems to the students—typically, 7 sonnets (6 by Shakespeare, one by de Vere) to each student or 24 sonnets (21 by Shakespeare, 3 by de Vere) to each group of three students. The students' collected data are combined into a single data set for further analysis through the use of graphing calculators or computer software. Each student is given an electronic copy of the Shakespeare data set and the de Vere data set; these data sets include the word length and the syllable frequencies for each of Shakespeare's 154 sonnets and each of de Vere's 25 known poems. For homework or during the next class period, students are asked to read the Johnson (1994) article and then complete tasks and answer short questions that encourage them to begin exploring the data. Several student prompts are provided:

1. Indicate the number of words in all 154 Shakespeare sonnets consisting of one letter, two letters, three letters, and so on. Repeat this procedure for the 25 de Vere poems.
2. Calculate the relative frequencies for words of all possible lengths for both the Shakespeare and the de Vere data.
3. Using two or three techniques of your choosing, compare the distributions of the word-length frequencies in Shakespeare's sonnets and in de Vere's poems.
4. On the basis of these calculations, do you think that Shakespeare was actually de Vere? Comment and include specific calculations in your response. Include graphs to strengthen your arguments.

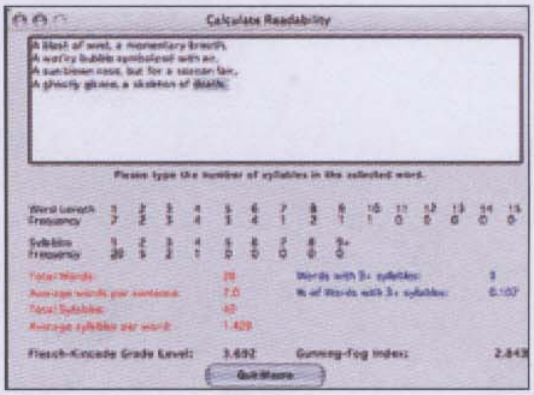



**Fig. 2** The macro provides readability statistics along with intermediate calculations.

5. Is it fair to compare Shakespeare to de Vere on the basis of the selections the class has analyzed? Note any concerns you have regarding our data or our methods of analysis.

## DATA ANALYSIS WITH TECHNOLOGY

### *Relative-Frequencies Histograms*

Students sum the word-length frequencies and enter these sums into the List Editor of their calculators (see **fig. 3**). Using the word-length frequencies found in de Vere's and Shakespeare's poems, students calculate the relative frequency of each word length for each author and then construct relative-frequency histograms. The steps required to construct these histograms are illustrated in **figures 4** and **5**.

As the plot in **figure 5** suggests, the relative frequencies of word lengths for the de Vere and the Shakespeare poems are quite similar (a numerical comparison of lists L4 and L5 shown in **fig. 4c** suggests the same). At first glance, this comparison appears to support the possibility that de Vere and Shakespeare are the same author. At a minimum, the analysis fails to show that they are different authors.

### *Box-and-Whisker Plots*

Although plots of relative frequencies are useful when examining the lengths of words that various writers use, they are not the only statistical tools students may use to study writings. **Figure 6** illustrates the use of box-and-whisker plots to compare word-length frequencies.

A closer inspection of the de Vere and the Shakespeare box-and-whisker plots (using TRACE features of the TI-84+) reveals that the median word length for both authors is identical (four letters). Likewise, first- and third-quartile word lengths are identical (three and five, respectively). Thus, the distribution of word lengths for de Vere and for Shakespeare appears to be quite similar.

### *Measures of Central Tendency*

Some students prefer to compare the writings of de Vere and Shakespeare by calculating the mean and the standard deviation of the length of words each author uses. **Figures 7** and **8** illustrate steps to calculate the mean word length for both authors.

The mean and standard deviation for word lengths for de Vere's poems (4.01 and 1.76 letters, respectively) are similar to those for Shakespeare's sonnets (4.15 and 1.91 letters, respectively). These measures provide further evidence that the lengths of words found in de Vere's poems are similar to those in Shakespeare's sonnets.

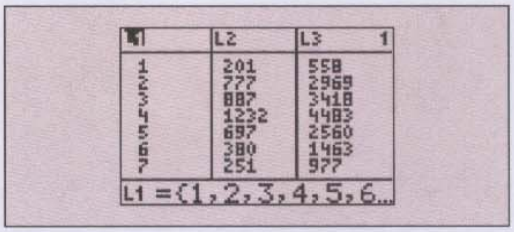



**Fig. 3** The number of letters in each word (list L1), the word-length frequencies found in de Vere's poems (list L2), and the word-length frequencies found in Shakespeare's sonnets (list L3) as they appear in TI-84+ List Editor

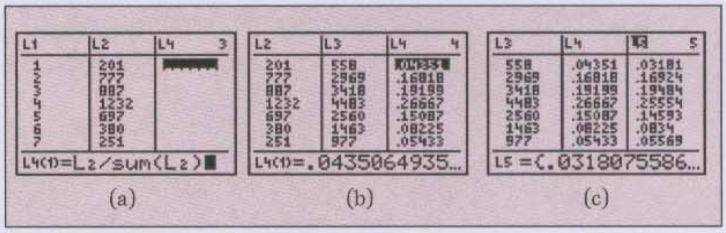



**Fig. 4** The formula for relative frequency for word lengths for de Vere is entered in list L4 (**a**); the relative frequencies for word lengths for de Vere and for Shakespeare are calculated in L4 (**b**) and L5 (**c**), respectively.

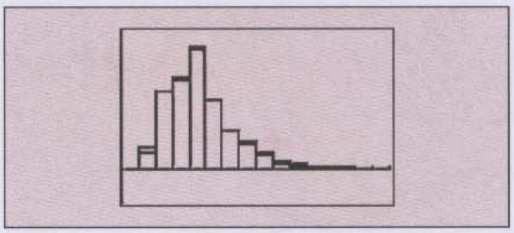


**Fig. 5** Simultaneous histograms of word-length frequencies for de Vere and Shakespeare

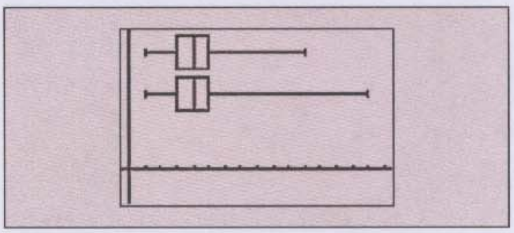


**Fig. 6** Simultaneous box-and-whisker plots of word-length frequencies for de Vere and Shakespeare

## CONCERNS REGARDING INITIAL DATA ANALYSIS

Regardless of the number of methods used to compare the de Vere and the Shakespeare data, not all students find similarities among word-length frequencies compelling. Typically, one or more students pose concerns such as the following:

1. The poems analyzed are primarily sonnets—a style of writing with very rigid structural

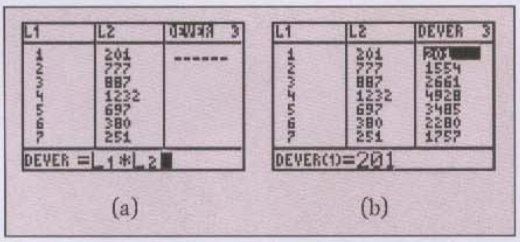



**Fig. 7** List DEVER is defined as the product of word length and frequency for de Vere's poems (**a**); total letters for each word length are stored in list DEVER (**b**).

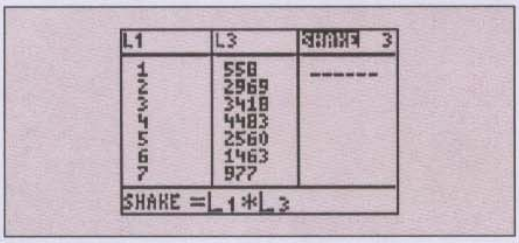



**Fig. 8** List SHAKE is defined as the product of word length and frequency for Shakespeare's sonnets.

requirements. For example, sonnets have fourteen lines, each line is written in iambic pentameter and has ten syllables per line, and the sonnet form devised by Shakespeare follows an *abab cdcd efef gg* rhyme scheme. Given the strict form of sonnets, it is plausible that all sonnets will have word lengths with relative frequencies similar to those of Shakespeare and de Vere.

2. Of the poems analyzed, all those written by Shakespeare were sonnets, whereas de Vere's poems were a mixture of styles. Comparing the Shakespearean selections to those attributed to de Vere is therefore akin to comparing apples with oranges. For comparisons to be valid, only sonnets should be analyzed.

To address the first concern, you may have students analyze one or more sonnets by authors other than Shakespeare or de Vere. Graphs of relative frequencies of word lengths for two sonnets—"Happy Ye Leaves! Whenas Those Lily Hands" by Edmund Spenser (ca. 1552–99) and "A Blast of Wind, A Momentary Breath" by Barnabe Barnes (ca. 1569–1609)—are shown in **figure 9**. These sonnets were chosen specifically because they were written by historically significant Elizabethan sonneteers.

Evidence such as that provided in **figure 9** suggests that relative frequencies of word lengths need not be the same for all sonnets, even for works composed by contemporaries of Shakespeare. Such findings seem to refute the notion that all sonnets have word lengths with relative frequencies similar to those of Shakespeare and de Vere.

To address the second concern—that investigations of authorship should be restricted to comparable modes of writing—you may have students reanalyze the de Vere data by including only the values taken from sonnets. Restricting our attention to a comparison of Shakespeare's sonnets with de Vere's sonnets does not dramatically affect previously drawn conclusions, namely, that relative frequencies of lengths of words used by Shakespeare and de Vere in the selected poems are quite similar.

### *Chi-Square Analysis*

To compare more rigorously the relative frequencies of the word lengths of Shakespeare's sonnets with those of de Vere's poems, students can use the chi-square statistic. Students begin by storing relative frequencies of word lengths for Shakespeare and de Vere sonnets into two lists.

Denoting the relative frequency of words of length $k$ in Shakespeare's sonnets as $s_k$ and the relative frequency of words of length $k$ in de Vere's sonnets as $d_k$, students calculate

$$\frac{(s_k - d_k)^2}{d_k},$$

the square of the deviation divided by the expected relative frequency for each $k$. Using a graphing calculator, students calculate

$$\frac{(s_k - d_k)^2}{d_k}$$

for $1 \le k \le 10$ in a single list (see **fig. 10a** and **10b**). The chi-square statistic equals the sum of the square of the deviation divided by the expected relative frequency for each $k$

$$x^2 = \sum_{k=1}^{10} \frac{(s_k - d_k)^2}{d_k} x$$

and can be calculated on a handheld graphing calculator by summing a single list (see **fig. 10c**).

In general, the larger the chi-square statistic is, the larger the differences between the relative frequencies are. To determine the likelihood that the lengths of words in Shakespeare's sonnets are distributed in a manner consistent with the lengths of words in de Vere's sonnets, students use the cumulative distribution function (cdf) on their calculators to determine the $p$-value associated with the chi-square statistic. The resulting $p$-value of 0.998 indicates that the distribution of word lengths in Shakespeare's sonnets is consistent with the distribution of word lengths in de Vere's work—in fact, we can expect that word lengths come from the

same distribution 99 times out of 100. Although this result does *not* tell us that Shakespeare and de Vere are the same person, it fails to provide us with evidence that would allow us safely to reject the hypothesis.

Students are reminded that the chi-square test and other statistical analyses provide only one form of evidence regarding Shakespeare's true identity. To make more compelling and conclusive arguments regarding the Bard's true identity, researchers (including my students) need to include other forms of evidence in their analyses. Such comments may be used to spur students toward continued interdisciplinary work in answering such questions.

## SUMMARY

Although our investigations have not provided my students with evidence sufficient to conclude that William Shakespeare was Edward de Vere, the project has encouraged them to reconsider mathematics as a useful, interesting field of study. Although my intention is not to reduce the work of Shakespeare (and others) to average word length and syllable count—the Bard's skillful use of wit, irony, allusion, and personification as well as his artful elaboration of timeless themes set his work apart—the statistical study of Shakespeare's writing has proved worthwhile for my students.

As my students analyze word length data with Microsoft Excel or their TI-84+ calculators, they use research methods similar to those university researchers use (Johnson 1994) and are often surprised to discover how accessible such techniques are. Moreover, my students have found the activities motivating because, as they construct histograms and box-and-whisker plots, they attempt to answer a question that has puzzled readers, scholars, and historians for generations: Who was the real William Shakespeare?


## ACKNOWLEDGMENTS

The author extends thanks to his colleagues in the Department of Teacher Education and the Howe Center for Writing Excellence at Miami University for support and guidance regarding this and other interdisciplinary initiatives.


## REFERENCES


Borasi, Raffaella, and Marjorie Siegel. *Reading Counts: Expanding the Role of Reading in Mathematics Classrooms*. New York: Teachers College Press, 2000.

Dawkins, Peter. *The Shakespeare Enigma*. London: Polair, 2004.

Johnson, Eric. "Comparing Texts and Identifying Authors." *TEXT Technology* 4, no. 1 (Spring 1994): 7–12.

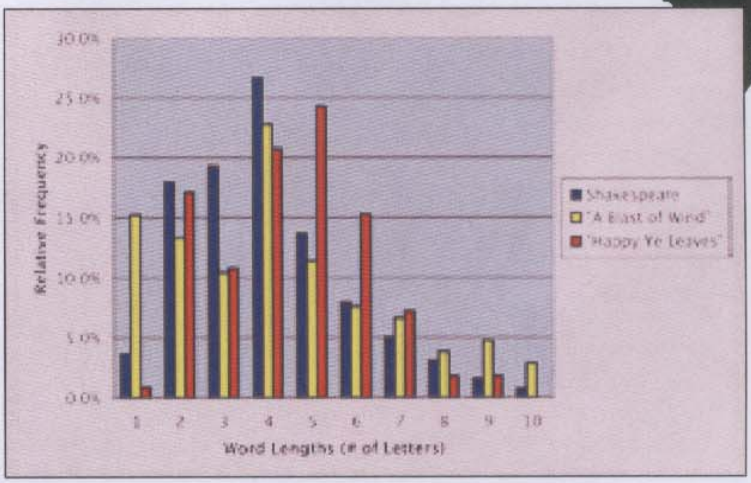




**Fig. 9** A bar graph comparing relative frequencies of word lengths for Shakespeare with those of other Elizabethan sonneteers

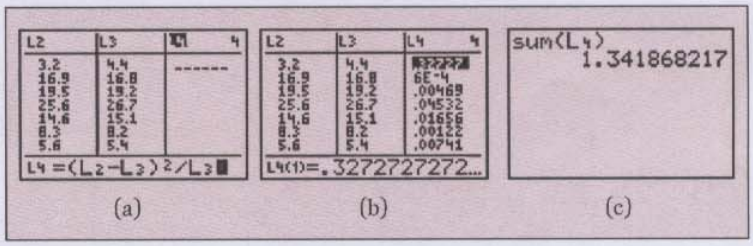



**Fig. 10** List L4 is defined as the squared deviation divided by expected relative frequency (**a**); list L4 is populated with results of this calculation for $1 \le k \le 10$ (**b**); chi-square statistic from List L4 (**c**)


Martinez, Joseph, and Nancy Martinez. *Reading and Writing to Learn Mathematics: A Guide and a Resource Book*. New York: Allyn and Bacon, 2001.

Mower, Pat. *Algebra Out Loud: Learning Mathematics through Reading and Writing Activities*. New York: Jossey-Bass, 2003.

Satchell, Michael. "Hunting for Good Will: Will the Real Shakespeare Please Stand Up?" *U.S. News and World Report*, July 24, 2000. www.usnews.com/usnews/culture/articles/000724/archive_014659.htm.

Shakespeare Fellowship. "The Poems of Edward de Vere." 2006. www.shakespearefellowship.org/virtualclassroom/deverepoems.htm. ∞



MICHAEL TODD EDWARDS, edward m2@muohio.edu, is an assistant professor of mathematics education in the Department of Teacher Education at Miami University in Oxford, Ohio. His research interests include the use of graphing calculators and dynamic geometry software in the teaching and learning of mathematics.