19. EXAMINING THE EDUCATIONAL POTENTIAL OF COMPUTER-BASED TECHNOLOGY IN STATISTICS

Peter Jones
Swinburne University of Technology

INTRODUCTION

By its very nature, statistics is a computationally intensive endeavor. The hand calculation of a correlation coefficient, such as the Pearson product moment correlation coefficient, is tedious, extremely time consuming, and has great potential for computational error. To assist with the task and reduce the potential for error, methods have been developed that reduce the computation to the completion of a carefully constructed table (e.g., Moore & McCabe, 1993). With the advent of statistical packages and, more recently, scientific calculators with statistical functions, students with routine access to such technology have no need to master such computational procedures. Yet textbooks are now being written that promote the use of spreadsheets to replicate such computations (Daly, Cody, & Watson, 1994), despite the fact that the same spreadsheet, once the data has been entered, will calculate the value of the required statistic on demand!

Why do such misuses of computer-based technology emerge? In part, at least, it is due to the general lack of recognition that statistics, like all human intellectual activity, has always been shaped by available technology; however, with time, the technologies "become so deeply a part of our consciousness that we do not notice them" (Pea, 1993, p. 53). There is a tendency for the technology determined activity to achieve an intellectual status in its own right. The construction of tables to aid with complex statistical computations is a product of a time when the only computational technology available to statisticians was pencil and paper, tables of logarithms, and, possibly, some form of mechanical adding machine. In that era, mastery of such computational techniques was a necessary prerequisite for further statistical analysis which, ultimately, was the goal. However, in the statistics classroom, the time and effort required for students to master these computational procedures and the time taken to complete even a single computation with any sort of real data meant that performing the computation often became an end in itself, and the process became valued in its own right as an intellectual achievement.

Thus, while the goals of introductory statistics courses tended to emphasize understanding and application, in practice assessment in statistics has been based on the mastery of computational procedures such as the calculation of a *t* test. This may have been necessary in the past and may be of marginal but possibly justifiable value in the present with the limited availability of computer-based technology in most classrooms. It will, however, be totally inappropriate when, in the not too distant future, students have powerful statistical software at their finger tips at all times, in much the same way as students in affluent countries have access to cheap statistical calculators today. What will be our goals then, and what role will computer-based technology play in achieving those goals?

To answer this question, we need to look beyond the view of computer-based technology as a means of enhancing the teaching and learning of current curricula; the end result of such activities is often no more than a translation of what are essentially pencil-and-paper-based activities onto a computer screen, albeit often done in an exciting and enlightening manner. As we move into an era in which computer-based technology becomes the new pencil and paper, such developments will become of historical interest at most (Kaput, 1992). Although there is undoubted benefit in using computer-based technology to reduce the time students spend on statistical computation, or in using it to illustrate the Central Limit Theorem, for example, the ultimate power in the technology lies in its ability to reshape the nature of intellectual activity in the statistics classroom. To see why this might be, we need to look generally at the ways in which interacting with technology of this sort has the potential to affect human intellectual performance. We will do this by using a theoretical framework proposed by Salamon, Perkins, and Globerson (1991) which has implications for both future classroom practice and research.

INTELLIGENT TECHNOLOGY

A key element in the theoretical framework we will use is the concept of intelligent technology; that is, “a technology that can undertake significant cognitive processing on behalf of the user” (Salamon, Perkins & Globerson, 1991, p. 3). At the lowest level, pencil and paper is an example of intelligent technology. For example, in statistics we might use pencil and paper as a memory-saving device to record a set of data values and then use it to record the steps in our hand calculation of a correlation coefficient. Without access to such technology, such computations would be beyond the intellectual capacity of all but a few, because having to remember the data values at the same time as mentally performing the calculation would almost invariably lead to cognitive overload. Even with pencil and paper, such calculations require a considerable intellectual effort on the part of the user and, in the past, have severely limited what can be achieved in the statistics classroom. What then differentiates the new computer based technologies from the old pencil and paper based technologies? In common with the older technologies, computer based technologies can act as a storage device for information. However, they also have the added dimension of being able to carry out significant processing of that information on behalf of the user, literally at the press of a button. Data can be entered into a statistical package or statistical calculator and then stored. Once entered, complex statistical calculations such as fitting a least squares line are routine.

Computer based technologies clearly have the potential to significantly support human intellectual performance. Salamon et al. (1991) have suggested that, in the classroom, it is useful to think of this support occurring in two very different ways. The first is when technology is used as a tool to amplify the skills of students. For example, technology may be used to enable students to carry out complex calculations that would be beyond their capabilities without the aid of the technology. The intellectual outcomes that arise in such circumstances are termed effects *with* the technology.

The second involves using the technology to create learning activities that might help students develop increased understanding or knowledge but that does not necessarily require them to have technological support to implement this understanding or knowledge. That is, working with the technology brings about lasting changes in the students’ cognitive capabilities. Intellectual outcomes that arise in such circumstances are termed effects of technology. For example, consider what might happen when a student uses computer-based statistical software to analyze bivariate data. The software will almost certainly increase the student's capacity to analyze the data by automating the process of, for example, constructing scatterplots, calculating Pearson's *r*, and fitting a least squares line to the data. This would be considered an effect of working with the technology. It might also be found that, as a result of using the software in carrying out this and other bivariate analyses, the student develops a deeper understanding and knowledge of the ideas of bivariate analysis in general. This would be termed an effect of the technology. To examine the consequences of these ideas for the teaching and learning of statistics, we will use the emerging technology of the graphics calculator.

THE GRAPHICS CALCULATOR

The distinction between a calculator and a computer is no longer clear. Brophy and Hannon (1985) stated that "in mathematics courses, computers offer an advantage over calculators in that they can express results graphically as well as numerically, thus providing a visual dimension to work with variables expressed numerically" (p. 61). This difference disappeared in 1986 when a programmable scientific calculator with interactive graphics signaled the emergence of the graphics calculator.

A graphics calculator's full functionality cannot be judged from its keyboard. Similar to a modern computer, it uses menus to guide the user through its many and varied capabilities. Of interest to statistics educators are the most recent graphics calculators (e.g., the TI-83) that have many of the capabilities of a relatively sophisticated computer-based statistics package. For example, the TI-83 has the following statistical capabilities:

- Spreadsheet-like data entry and modification.
- Statistical graphics: scatterplots, line graphs, boxplots (with and without outliers), histograms, and normal probability plots.
- Descriptive statistics: univariate and bivariate.
- Regression models: median-median, linear, quadratic, cubic, quartic, logarithmic, exponential, power, logistic, and sinusoidal.
- Inferential statistics: *z* and *t* -tests (one and two sample), tests for proportions (one and two sample), two-sample *F*-test, linear regression; and simple one-way ANOVA, all with associated confidence intervals and distribution functions available in both numerical and graphical form.

The graphics calculator also has the capability of interfacing with a computer or another graphics calculator for data exchange. It does this for a fraction of the cost of buying a computer to run the equivalent software; for example, for the cost of a single software-equipped computer, a school could purchase a set of 20 graphics calculators, each of which has much the same statistical capability. Thus, the potential exists, at least in more affluent countries, for powerful statistical computing to be at the finger tips of students at all times. This is a qualitatively different situation than what we have had to date, when students could only be assumed to have limited access to such technology. What sort of educational experiences will be available to students in possession of such calculators in the statistics classroom?

WORKING WITH TECHNOLOGY: THE INTELLIGENT PARTNERSHIP

When using computer-based technology such as the graphics calculator to conduct a statistical analysis, the potential exists for the formation of what Salamon and his colleagues have termed an "intelligent partnership" (Salamon et al., 1991, p. 4). Potentially, such partnerships can lead to a level of statistical performance that would not be possible without the technology. This goes beyond simply enabling the student to carry out more difficult computations. For a partnership with technology to be "intelligent" there must be a complementary division of labor between the user and the technology, and the user of the technology must be mindfully involved in the task (Salomon & Globerson, 1987).

Suppose a student with access to a graphics calculator such as the TI-83 is given data on the weights and girths of a number of trees (in this case n = 99) and is asked to investigate the relationship between the two variables. The real world purpose of the exercise is to be able to predict a difficult quantity to measure in the forest, such as the weight of a tree, from an easy quantity to measure, such as the tree's girth. The data can be either entered into the graphics calculator manually or, if already stored in electronic form, transferred electronically. The graphics calculator stores the data in columns (called "lists" in the TI-83) as is common to most modern computer packages; weights are stored in List 1 (L1) and girths in List 2 (L2) (see Figure 1).

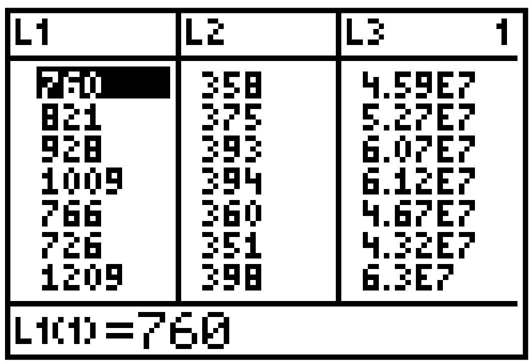
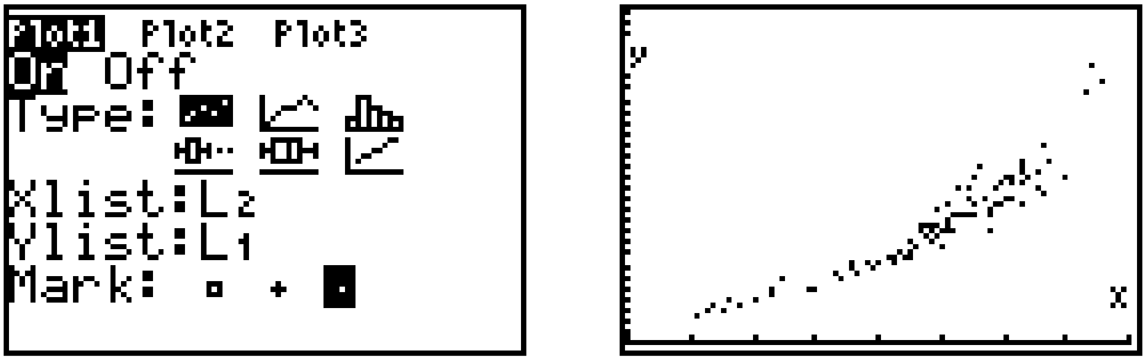


Figure 1: Graphics calculator display showing how the data are entered

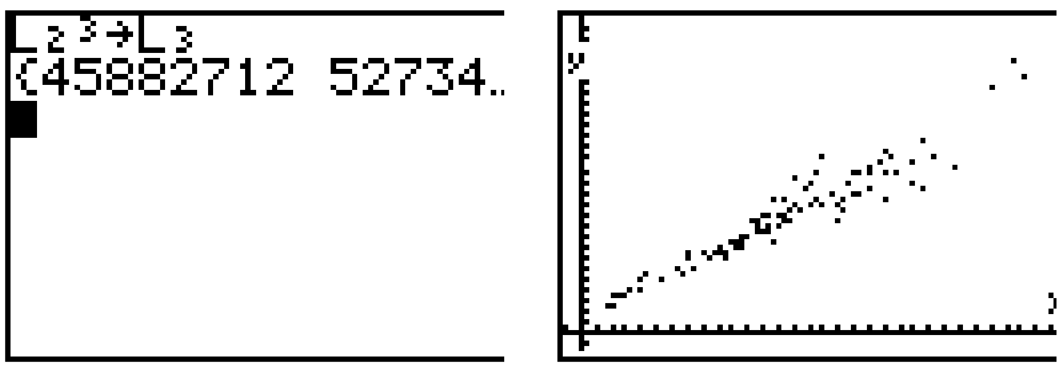
Following good statistical practice (e.g., Moore & McCabe, 1993), students begin with a graphical analysis to see whether or not the two variables appear to be related and, if so, whether or not the relationship can be assumed to be linear. Using their knowledge of bivariate statistics, they decide that the graphical display that is appropriate in this situation is a scatterplot. The responsibility for constructing the scatterplot can then be passed over to the graphics calculator. However, because the graphical analysis is a precursor to a regression analysis, students need to consider which variable is the predictor variable, because this will be plotted on the horizontal (X) axis. In this case, tree girth will be used to predict tree weight; thus, tree girth is entered into the calculator as the predictor variable (X) (see Figure 2a). The resulting scatterplot is shown in Figure 2b.

From their knowledge and understanding of scatterplots, students should realize that the scatterplot is consistent with a moderately strong positive relationship between tree weight and tree girth, but that the relationship is clearly nonlinear. Using an appropriate data transformation, however, it is likely that the relationship can be linearized. The computational aspects of this task would be completed by the calculator, although a number of transformations might be proposed and tested before an appropriate one is obtained. The result of the linearization process, using the transformation $X^3 \rightarrow X$ is shown in Figure 3.



2a. **2b.**

Figure 2: Graphic calculator displays showing the setup for constructing a scatterplot (2a) and the resulting scatterplot (2b)

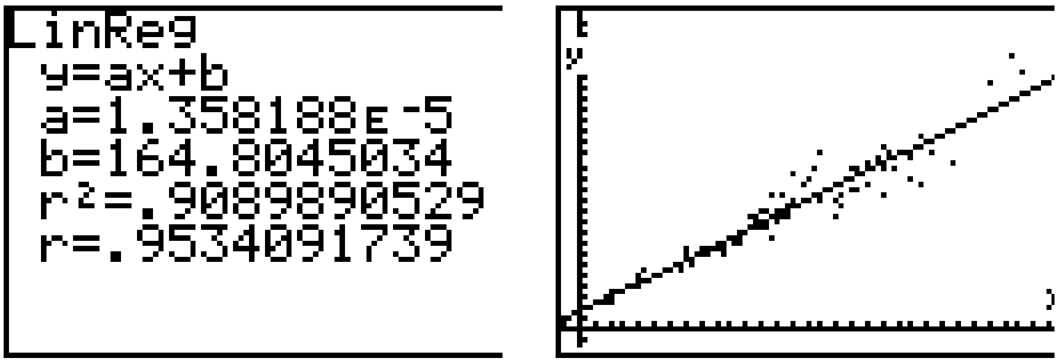


3a. **3b.**

Figure 3: Graphic calculator displays showing the setup for transforming data (3a) and the resulting scatterplot (3b)

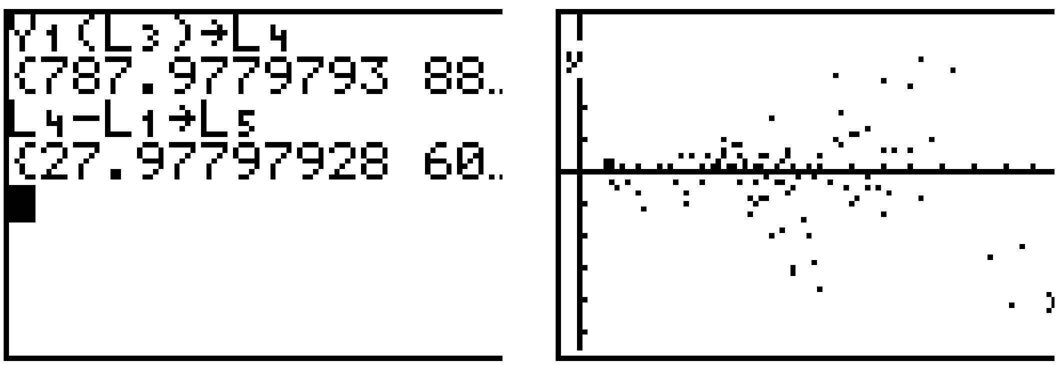
Once the linearization procedure has been conducted to the satisfaction of the users, the graphics calculator will find the equation for the least squares line of best fit and display this line on a scatterplot (see Figure 4). The student then has the responsibility of interpreting the information generated by the technology and translating the information back into the language of the specific problem Remembering that X, the calculator predictor variable, now represents girth cubed the regression equation relating the two variables is: weight = 164.8 +.0000136 girth cubed.

The goodness-of-fit of the cubic model can be checked graphically by using the calculator to obtain a residual plot (see Figure 5). However, it is up to students to interpret and make judgments about the residual plot. The coefficient of determination is also available (Figure 4a), which can be used to quantify the strength of the relationship. Again, it is up to students to interpret the results. In this case, $R^2 = r^2 = .91$, so that 91% of the variation in tree weights can be explained by the cubic regression model.



4a. 4b.

Figure 4: Graphic calculator displays showing regression output (4a) and scatterplot displaying the relationship between tree weight and girth with the least squares line of best fit displayed (4b)



5a. 5b.

Figure 5: Graphical calculator displays showing calculation of residuals (although this can also be done automatically) 5a and resulting residual plot 5b

The partnership between students and the graphics calculator just described is intelligent in that students are mindfully engaged in the activity, and there is a complementary division of labor. In this partnership, students take responsibility for planning and implementing the analysis but, at the appropriate times, pass responsibility over to the technology to, for example, construct the scatterplot, or carry out the computations required to linearize the data. A crucial aspect of the partnership is the constant monitoring and checking of the information generated by the calculator to make sure that the solution produced is consistent with the users' knowledge and understanding of the statistical techniques being used and the specific problem.

With an intelligent partnership, the potential exists for the combination of the student and calculator to be “far more ‘intelligent’ than the human alone” (Salamon et al., 1991, p. 4). For example, with access to technology such as the graphics calculator, students have the potential to develop skills at analyzing bivariate data, as illustrated above, that greatly exceeds what they could ever hope to achieve using pencil and paper alone. Unfortunately, these intelligent partnerships do not appear to be self-generating, and the challenge for teachers is to develop instructional strategies that promote their formation. It is also unlikely that they will be realized unless students have routine access to the necessary technology, just as professional statisticians have routine access to their computers. Finally, there is a need to reassess what is taught, because the knowledge and understandings needed to develop an intelligent statistical partnership when working with technology are almost certain to differ in some significant ways from those needed for students who will compute statistics without access to technology.

The possibility of students forming intelligent partnerships with technology in statistics gives them the potential to work at a level in statistics that may be totally unachievable without the technology. This in effect calls into question our traditional notions about what constitutes statistical intelligence and how it should be assessed. Should it be measured by the statistical performance of the student working without any technological aid, or does the possibility now arise of it being also recognized as the statistical performance of a joint system? If we accept that a student working in an intelligent partnership with computer-based technology is a legitimate and valued form of statistical activity, then we must consider the possibility that appropriate assessment of statistical intelligence involves assessment of that partnership. Further, given that in the long run almost all real world statistical activity involves the use of some supportive computer-based technology, it could be argued that one of our prime pedagogic interests in statistics should be directed at the task of developing instructional strategies for building and assessing the statistical intelligence of such partnerships and not just the individual acting in isolation.

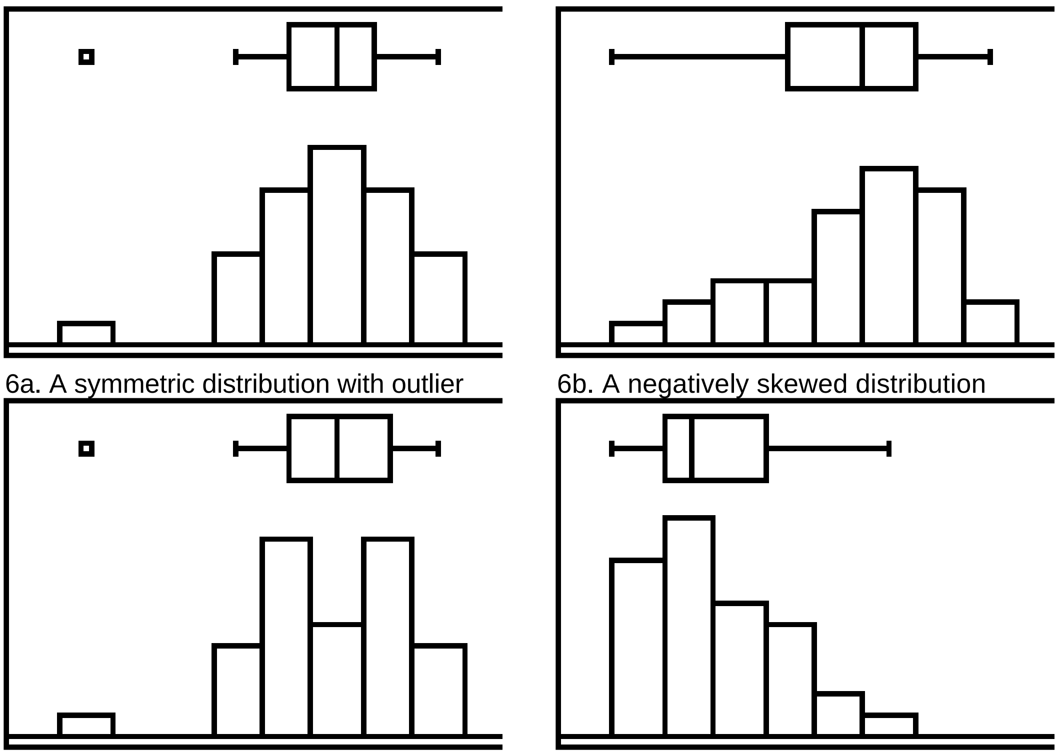
DEVELOPING UNDERSTANDING: POTENTIAL EFFECTS OF LEARNING STATISTICS WITH TECHNOLOGY

The use of computer-based technology to provide learning experiences to help build understanding of the concepts and ideas that underlie statistical theory has long been promoted by statistics educators as a means of enhancing the teaching and learning of statistics (e.g., Bloom, Comber & Cross, 1985; Thomas, 1984). In Salamon et al.'s (1991) terms, this is an effect *of* the technology. When technology is used in this way in the teaching and learning of statistics, its prime purpose is to provide a learning experience that will develop statistical understandings and insights rather than just generate statistical results, although these may be a by-product of the activity.

For example, boxplots provide a visually powerful and succinct method for displaying the essential features of a data set and are frequently used in statistical analysis. In data analysis, having generated a boxplot, it is a useful skill to be able to picture the general form of the dataset from which it was derived, in particular the symmetry of the data distribution and the location of any potential outliers. Using the ability of a graphics calculator such as the TI-83 to overlay its statistical plots, we can provide a learning experience for students whose primary purpose is to build these connections. Using carefully selected datasets, histograms chosen to illustrate particular aspects of data distributions can be displayed simultaneously with their associated box plots so that students can see the relationship between the general form of a histogram and its corresponding boxplot (see Figure 6).

Figures 6a and 6c show quite clearly the presence of a potential outlier in the histogram which can be seen to correspond to the isolated point in the boxplot. The parts of the histograms representing the remainder of the data are symmetric in both cases, and this is reflected in the symmetry of the corresponding boxplots. Note that one distribution is bimodal, but this is not reflected in the box plot, which shows the relative insensitivity of box plots to variations in the center of a distribution. Figures 6b and 6d show how the asymmetry in the histograms is reflected in the corresponding boxplots.

The achievement of long-term effects of technology on students' understanding rests on the fundamental assumption that "higher order thinking skills that are either activated during an activity with an intellectual tool or are explicitly modeled by it can develop and become transformed to other or at least similar situations" (Salamon et al. 1991, p. 6). In this particular exercise, the expectation is that using a graphics calculator to simultaneously display a range of histograms with their associated boxplots in a variety of situations will help improve students' interpretive skills. Unfortunately, while this appears to be a perfectly reasonable assumption, at present little research has been conducted concerning the effect on understanding of working with intelligent technology and the type of technology based instructional sequences that support the development of this understanding.



6a. A symmetric distribution with outlier

6b. A negatively skewed distribution

6c. A symmetric distribution (bimodal) with outlier

6d. A positively skewed distribution

Figure 6: Selected histograms with corresponding boxplots simultaneously displayed

DISCUSSION AND CONCLUSION

This paper re-examined the potential of computer-based technology in statistics from the perspective of students' possible interactions with it to help them calculate and learn statistics. The graphics calculator was chosen for its potential to provide students with access to powerful statistical software at a fraction of the cost of the equivalent computer-based system. However, the conclusions drawn are hardware independent. Using a theoretical framework (Salamon et al., 1991), we distinguished between the potential effect of computer-based technology on statistical intelligence when a student is working in partnership with the technology and the effect on statistical intelligence when the student is working without the aid of technology but where the technology has been used to help support instruction. Given that most real world statistical activity will be conducted with the aid of computer-based technology, increasing emphasis must be given to building the sort of user/technology partnerships that will lead to optimizing the statistical intelligence of the *partnership* rather than the individual. In doing this, we should also be aware that the skills developed will not necessarily be the same as those required to improve competence in traditional pencil and paper based statistics.

Further, there is no reason to assume that the technology that is best for optimizing the statistical performance of a user/technology partnership is the best technology for developing conceptual understanding. Different learning goals require different instructional strategies and different technological support. When using computer-based technology in an intelligent partnership, the student is concerned with the solution of a *particular* problem. When using computer-based technology to develop understanding, the particular solution is of little educational consequence--what is necessary for the acquisition of statistical understanding is technological support that facilitates *abstraction* from the particular to the general.

In conclusion, although computer-based technology is not yet an integral part of the statistics classroom, we must start preparing ourselves for the time when this will be the case. This means looking beyond the view of what constitutes statistical intelligence when a student is using primarily pencil-and-paper activities, although possibly supported by some basic computational aid, to a view that recognizes that, for much of the time, the student will be working in a computer environment. In this environment, it is possible to give statistical intelligence a new meaning by recognizing the performance enhancing potential of the technology when used in an appropriate manner. If this partnership view of statistical intelligence is to be developed, time and energy must be used to identify and develop the knowledge and skills necessary to help our students build these partnerships.

Finally, the instructional strategies we use with computer-based technology will depend critically on whether our goal is statistical competence with the technology or enhanced statistical understanding as a result of having worked with the technology. To date there seems to have been a confusion of aims and this, coupled with the relatively short time that we have been working with the technology, explains our limited success with computer-based technology in the classroom.

REFERENCES

Bloom, L. M., Comber, G. A., & Cross, J. M. (1985). Using the microcomputer to simulate the binomial distribution and to illustrate the central limit theorem. *International Journal of Mathematics Education in Science and Technology, 17*, 229-237.

Brophy, J., & Hannon, P. (1985). On the future of microcomputers in the classroom. *The Journal of Mathematical Behaviour, 4*, 47-67.

Daly, T., Cody, M., & Watson, R. (1994). *Further mathematics*. Sydney: McGraw-Hill.

Kaput, J. J. (1992). Technology and mathematics education. In D. A. Grouws (Ed.), *Handbook of research in mathematics teaching and learning* (pp. 515-556). New York: MacMillan.

Moore, D. S., & McCabe, G. P. (1993). *Introduction to the practice of statistics* (2nd ed.). New York: W. H. Freeman and Company.

Pea, R. (1993). Practices of distributed intelligences and designs for education. In G. Salamon (Ed.), *Distributed cognitions: Psychological and educational considerations*. Cambridge University Press: New York.

Salomon, G., & Globerson, T. (1987). Skill may not be enough: The role of mindfulness in learning and transfer. *International Journal of Educational Research, 11*, 623-627.

Salamon, G., Perkins, D. N., & Globerson, T. (1991). Partners in cognition: Human intelligence with intelligent technologies. *Educational Researcher, 20* (3), 2-9.

Thomas, D. A. (1984). Understanding the central limit theorem. *Mathematics Teacher, 77,* 542-543.