Active learning of statistics: a case study

Erica J. Morris and Eileen Scanlon
Institute of Educational Technology, Open University
email: e.j.morris@ope.ac.uk

Research at the Open University has investigated students' learning of statistical concepts and how information technology can be effectively used to support this process. Previous empirical work has looked at psychology students' misconceptions relating to correlation and how computer-based learning environments can be used to address these. This paper reports on the findings from a qualitative study that investigated students' learning collaboratively from a multimedia application called ActivStats.

Introduction

Students studying psychology or a social science have to take courses in statistics, but students often experience difficulties in learning statistical concepts and techniques. Indeed, there is a body of interdisciplinary research that indicates that students tend to have confusions and difficulties with a variety of topics in statistics (Garfield and Ahlgren, 1988; Shaughnessy, 1992). It has been found that students hold misconceptions relating to fundamental statistical concepts, such as the mean and correlation (Mevarech, 1983; Morris, 1997). Research in computer-based learning has looked at how the computer can be used effectively to address students' statistical misconceptions and help students come to understand key statistical ideas (Cumming and Thomason, 1995; Morris, 1998).

It has been argued that the design of computer-based learning materials should exploit the capabilities of computer technology (Shuell, 1992; Milheim, 1996). For example, in contrast to traditional forms of instruction, such as text, the computer can provide the learner with multiple representations that are dynamically linked and offer immediate feedback to their actions at the human-computer interface. In the case of statistics, computer technology can be used to offer multiple representations of statistical concepts that are computationally linked so that changes in one representation respond to changes in another. For example, computer-based learning materials can be designed so that when a learner changes the available data set, or the value of a statistic, such as a correlation coefficient, they will see the corresponding change in the data displayed on a scatter plot.

Research at the Open University has been looking at how computer-based learning can be used effectively to support students' learning of statistics. Previous work in this area concerned the development and evaluation of a computer-based learning system called *Link* (Morris, 1998).

Link

Link is a computer-based learning system that was implemented at the Open University and was designed to be used by psychology students to review their understanding of correlation (Morris, 1998). The design of this system was informed by research-based principles of learning, a review of computer-assisted learning programs covering correlation, and formative and expert evaluation studies. Unlike the systems that were reviewed, *Link* makes use of data from authentic studies in psychology and provides learner activities that were specifically designed to address students' misconceptions about correlation. For example, empirical work has shown that students can hold misconceptions concerning negative correlations and the strength of correlations: students think that a negative correlation indicates no relationship between variables and/or that a positive correlation is stronger than a negative correlation when this is not the case. Learner activities in the system were therefore designed to address these confusions. For example, activity 2 was designed to remedy a student's unidirectional conception of correlation where they inappropriately think that a negative correlation indicates no relationship (Figure 1). In this activity a student must decide from a set of correlations the coefficient that represents the scatter plot that is displayed. Each of the correlations in the data table are linked to a graph so that when a user selects a correlation a scatter plot that represents the relationship is presented. A user is therefore provided with graphical feedback and text-based feedback that describes the relationship represented by the correlation that was selected.

Figure 1: A learner activity in Link
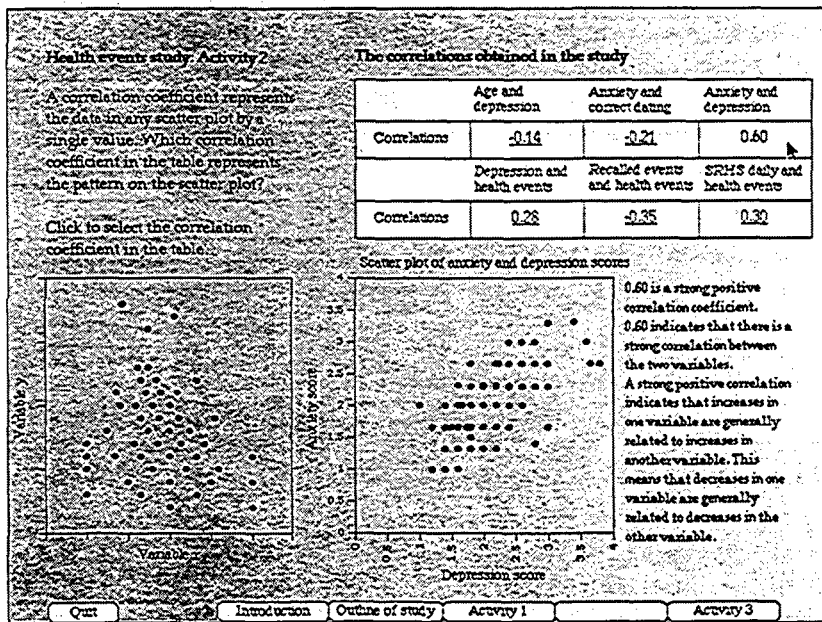
A summative evaluation study of *Link* was undertaken involving fifty psychology students to assess the effect on students' understanding of correlation. It was found that the use of *Link* significantly contributed to students' general understanding of correlation (Morris, forthcoming). Although this empirical work indicated that the use of *Link* contributed to students' understanding of correlation, one of the limitations of the system relates to the design of the learner activity described above. Due to the constraints of using genuine data sets, the learner is not able directly to modify the data or the data points on the scatter plot and see the corresponding changes in a linked representation. This suggested that further work was required to look carefully at whether students' understanding of correlation is affected by the use of learner activities that provide multiple representations that are computationally linked.

The study described in this paper was conducted to investigate students using a multimedia learning resource called *ActivStats* (*ActivStats*, 1997) and whether particular kinds of learner activities affected the learning process.

ActivStats

ActivStats is a multimedia learning resource that also provides learner activities for correlation. The study described in this paper was designed to investigate whether these activities might affect students' understanding of correlation and to identify aspects of *ActivStats* that students find valuable in terms of the learning process.

The study involved pairs of students who were asked to work together on a series of activities in *ActivStats*. The students were asked to talk between themselves about what they were thinking and doing while they used the resource, which meant that qualitative data concerning the process of learning could be gathered. Issroff (1993) has described a methodology for research in computer-supported cooperative learning which has provided practical guidelines to help researchers to design and study effective computer-supported co-operative learning situations. Accordingly, these guidelines were followed as appropriate. For example, Issroff (1993) has recommended that students should be told that they are expected to work together on the computer task and that studies should involve friendship groupings because these might be contributing factors to the success of the learning process.

Method

Participants
Six students who were taking psychology at a university took part in the study. The mean age for this group of students was 32 years, with the youngest being 20 and the eldest being 52. One of the students was studying for a single honours degree in psychology and five of the students were taking a joint degree that included psychology. Although four of the students were in the first year of their degree and two of the students were in their final year, all of the students had taken a course called 'Quantitative Methods for Psychologists' that covered correlation.

When students were asked to participate in the study they were asked to work with a friend or someone they felt happy to work with. Three pairs of students used *ActivStats* and the names of these students have been changed to assure anonymity: first pair (Mark and Trudy), second pair (Daniel and Alan), and third pair (Anna and Mary).

Software

ActivStats is a multimedia resource for learning statistics which uses the metaphor of a 'Lesson Book' to provide video, text, graphics and learner activities. In using *ActivStats*, students were asked to complete activities that included:

Observe how the magnitude of the Correlation Coefficient reflects the strength of the Linear Association. Answer the questions that are presented with this activity (Figure 2). Construct scatter plots with a given correlation. Answer the study questions that are presented with this activity (Figure 3 and Figure 4).

Figure 2: Observe how the magnitude of the correlation reflects the strength of the association

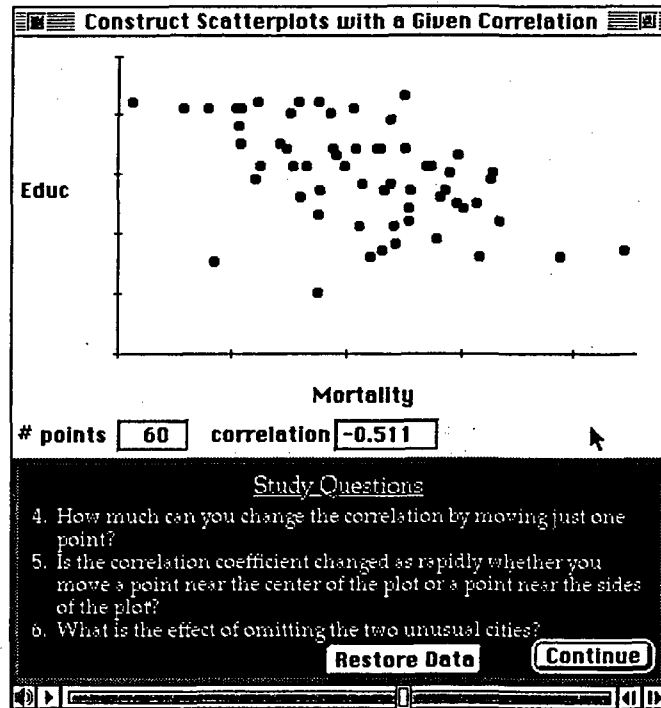*Figure 3: Construct scatter plots with a given correlation*

Figure 4: A relationship between mortality and education

Procedure

At the beginning of the session, students were given a set of instructions which told them that during the session they were to use the application *ActivStats*. They were asked to work together by completing specified sections of this application. Students were asked to talk between themselves about what they were thinking and doing while they worked through the program. Each pair of students was observed while they worked with *ActivStats* and by using a recording schedule a written and audio record of the session was made. Here, student-computer interactions and students' comments were noted.

When students had completed the specified tasks, they were interviewed and questions about the application were asked. For example, students were asked what they did and did not like about using *ActivStats* and what they thought about particular learner activities. Based on previous research (Morris, 1997), questions were also employed to look at the students' understanding of correlation:

(i) Which correlation is stronger 0.70 or -0.80?

(ii) Please arrange the set of correlations from the strongest to weakest relationship. (For this question, students were given five correlation coefficients on five small pieces of paper. So that the students could arrange the correlations in an order they were also given a piece of paper that provided a scale with 'Strongest' at one end and 'Weakest' at the other. The students could then place the correlations along this scale.)

Findings

The learner activities in ActivStats

Observe how the magnitude of the Correlation Coefficient reflects the strength of the Linear Association.

In this activity learners are provided with a scatter plot that is linked to a scale. This scale is designed to represent the correlation coefficient, r, and although it runs from -2.0 to 1.5, the slider is bounded at -1.0 and 1.0. Students can adjust the correlation coefficient by moving the slider to the left or right and can therefore observe the effect that this has on the scatter plot. This activity also presents two questions that the students were asked to answer: 'What does a correlation of 0 look like?' and 'How about a correlation of 0.5?' (Figure 2).

The first pair of students who worked together on this activity (Mark and Trudy) moved the slider right and left and observed the changes in the scatter plot. They talked while they did this:

Mark: 'Oh that's clever . . . I like that . . . You're doing a negative correlation now, aren't you? . . . So if you go the other way . . . should go to positive shouldn't it?'

Trudy: 'This is no correlation, right? Zero . . . Okay . . . This is positive.'

Mark: 'So we'll start going towards the positive correlation. So if you go right to 1 that should be nearly a straight line, shouldn't it? Forty five degrees (45°) . . . Yeah.'

This pair of students then went on to talk about the activity: Mark commented 'That is very good' and Trudy remarked, 'It's very visual.' Mark went on to comment:

'That's excellent . . . because . . . it's a hands on thing isn't it. You're actually doing it.'

The second pair of students who completed this activity also talked while they moved the slider and viewed the corresponding changes in the scatter plot (Daniel and Alan). While these students worked with this activity, they adjusted the slider and Daniel commented 'no relationship at all . . . negative correlation.' As they carried on with the activity, they read and answered the questions that are presented with the activity:

Daniel: 'What does a correlation of zero look like?'

Alan: 'A mess.'

Daniel: 'Yeap.'

Alan: 'How about a correlation of 0.5?'

Daniel: 'Sort of half way there.'

When the third pair of students (Mary and Anna) tried to answer the questions that are presented, Mary commented:

'Zero scattered everywhere but 0.5 scattered closer (to) each other. Like linear. [She laughs]. Sorry, I don't know.'

Construct scatter plots with a given correlation

This activity which involved students constructing a series of scatter plots, resulted in a number of interesting student-computer interactions. In addition, while completing this activity the students tended to take turns at constructing the scatter plots and worked together by commenting on their actions and the subsequent outcomes (Figure 3 and Figure 4).

On this activity students are given the opportunity to construct scatter plots with different correlations. To do this, students can create a scatter plot by clicking with the mouse to place data points on the graph and they are then able to drag the points to obtain a particular value of a correlation. Firstly, students are asked to construct scatter plots when they can simultaneously view the value of the correlation (Figure 3). Specifically, students are asked to build a scatter plot with a correlation of 0.5 and then one of -0.8. Secondly, students are shown a scatter plot of the age adjusted mortality and median education for 60 US cities and they are asked to move the points on this scatter plot to observe the changes in the correlation. For this part of the activity, students are asked specific questions including 'What is the effect of omitting the 2 unusual cities?' (Figure 4). Thirdly, students are asked to construct scatter plots of 0.5 and -0.8 in turn without seeing the value of the correlation. When they have constructed the scatter plot in both cases, they can then select the button 'Continue' to obtain the value of the correlation.

When Mark and Trudy attempted to construct a scatter plot of 0.5, Mark initially advised, 'So, forty five (45°) is a . . . 1' and they plotted five data points on the scatter plot to obtain as Trudy observed, '0.505 . . . Pretty close.' They then dragged the points to obtain an exact correlation of 0.5, and with reference to the constructed scatter plot, Mark then observed: 'A correlation of 1 is [a] forty five degree (45°) line . . . You can see a general pattern and also that's positive.'

Interestingly, Mark suggested that they try and obtain a correlation of -0.5. Mark did obtain a correlation of -0.5 and Trudy responded to this success by clapping. Mark then began to manipulate a data point on the scatter plot and the pair went on to comment:

Mark 'Only moving one point that's interesting . . . moving that one digit and it altered the whole process that's interesting.'

Trudy: 'Maybe it really depends on size of the sample.'

Mark: 'Yeah . . . But the general trend is there isn't it? . . . you've got the negative line running through there if you cast your imaginary line through it.'

Trudy: '. . . That means correlation is really sensitive.'

Mark: '. . . Yeah, so if you've got one of these silly outliers . . . it throws the stats completely out the window . . . stats, stats and damn lies.'

Trudy: 'Okay, let's continue.'

While completing the activity, Mark attempted to construct a scatter plot with a correlation of 0.5 without seeing the value of the correlation. In such a case, when a scatter plot has been constructed by the student, they can select the button 'Continue' and the program provides the actual correlation. Having selected 'Continue', Mark found that he had, however, constructed a scatter plot with a correlation of 0.9 and he commented, 'That was actually close to 1. You really want a lower, flatter curve.' Mark and Trudy then went on to try and construct a scatter plot with a correlation of -0.8 without seeing the value of the correlation. At first, Mark obtained a correlation of -0.1, the pair laughed and he commented, 'that was just a random guess that was'. Trudy then tried to obtain a correlation of -0.8 and found that she had constructed a scatter plot with a correlation of -0.7. In response to this, Mark observed, 'that's pretty good'. He then went on to comment:

'I'll give one more go just for a laugh . . . I'll try and see if I can do a linear one.' Mark then had a second attempt at a correlation of -0.8 and subsequently found that he had constructed a scatter plot with a correlation of -0.9. In the later stages of completing the activity, Mark remarked: 'It's a good learning tool to be able to play around with it.'

Out of the second pair of students, Daniel had a first go at constructing a scatter plot with a correlation of 0.5. Although Daniel only used four data points for this, he obtained the exact correlation of 0.5. To this, the pair of students laughed and Alan responded 'Perfect leave it there . . . got to do some more now.'

As described above, this activity provides an example relationship between education and mortality for 60 US cities. For this part of the activity, Alan moved various data points on the scatter plot and the value of the correlation changed accordingly. This part of the activity presents the question, 'How much can you change the correlation by moving just one point?' In response to this question, the pair of students moved data points on the scatter plot and Daniel decided that 'an outlier will change the value more.' The following question is also presented with this part of the activity: 'Is the correlation coefficient changed as rapidly whether you move a point near the centre of the plot or a point near the sides of the plot?' Daniel and Alan therefore moved a number of data points and Alan concluded, 'So moving one from the centre has more of an effect'.

The third pair of students, Anna and Mary, also took turns to try and construct scatter plots with different correlations. However, this pair did not try to drag the data points that they placed on the scatter plot. For example, before these students attempted to construct a scatter plot with a correlation of 0.5 for the second time, Mary asked Anna 'Can we use the points after we've placed?' and Anna went on to say, 'No, we have to answer the question'.

It is clear from the above findings that while students completed the activity they readily took turns in constructing scatter plots to obtain specified values of correlation. In addition, the recorded interactions between students and the comments that they made indicate that students helped each other by suggesting possible strategies to obtain a particular correlation. For example, when Daniel attempted to construct a scatter plot with a correlation of 0.5, Alan commented, '[Be] careful not to do too straight [a] line.'

The findings suggest that the above activity was valuable to students because it showed how the correlation coefficient represents data displayed on a scatter plot. The learner activity also enabled the first and second pair of students to see the effect that outliers have on the value of a correlation: the students dragged particular data points on a scatter plot and watched how the correlation changed.

Students' opinions of the software

All of the students were interviewed as a pair after they had used *ActivStats*. They were asked 'What did you like about using *ActivStats*?' In response to this question, Mark commented as follows:

> I would say that the hands on experience of being able to manipulate the data on the screen . . . you could alter it by playing around with it. It gave you probably a little bit better understanding of what was . . . meant by the figures 'cos so many people look at figures and it doesn't really mean anything whereas I think when we were taught in our stats course . . . to always look at graphs . . . you know as well as the figures. Then you

can see . . . the relationship . . . the interesting point about . . . statistics to me that they don't always tell the truth . . . and especially when you get things like outliers . . . you get that one little figure there and if you push him right on the outside it really can muck the sort of correlation up . . . You can visually do it . . . so that's the positive [aspect of the program].

Daniel also thought that the use of scatter plots in the program was helpful: 'I thought . . . the actual visual side of plotting your own scatter plot was very useful . . . Far better than wading through a textbook.' Indeed, Alan commented: 'Anything that isn't a text book is always going to be more interesting.'

Students were asked what they thought about the activity that provided a scale of a correlation that was linked to a scatter plot. To this Alan remarked, 'That was good. It showed it quite nicely', and Daniel replied, 'Yeah, I would agree with that definitely'.

Students were asked about the activity that involved constructing scatter plots with different correlations. Mark's comments concerning this activity were as follows:

> Again you know the visual impact was good . . . I was probably being very linear because I knew that if you go from the corner up to the other corner forty-five degrees that was a 1 on the correlation scale . . . so when they said sort of 0.8 I was trying to sort of drop it down and have a linear thing. But I realised that that's not necessarily the case. You can still have the plots scattered about . . . eventually I suppose it's an average isn't it? It's a mean, eventually, where it gives you the correlation figure. So that was a lesson to me if you like.

Daniel thought that the activity was 'good fun' and Anna commented, 'think it's a very good practical. I think it helps to understand.'

Students' understanding of particular concepts concerning correlation were also considered when students were interviewed. Students were asked which of two correlations, 0.70 or -0.80, was stronger. Two of the pairs of students (Mark and Trudy, and Daniel and Alan) were correct in their agreement that the stronger correlation was -0.8. However, of the third pair of students, Anna said that 0.7 was stronger and Mary said that -0.8 was stronger. They did not, however, consider or resolve this discrepancy.

Students were asked to arrange a set of correlations from the strongest to weakest relationship. The pair Mark and Trudy and the pair Daniel and Alan correctly arranged the correlations from the strongest to weakest relationship. Here, Mark's comments were appropriate: 'We just ignored the signs.' However, Mary said that she was confused when she tried to complete the task with Anna, and Mary said that she 'Can't decide'. In spite of this, this pair of students did correctly arrange the set of correlations.

Conclusions

The study described in this paper was designed to investigate whether learner activities in *ActivStats* might affect students' understanding of correlation and to identify aspects of this resource that students find valuable in terms of the learning process.

Computer technology can be used to provide the student with multiple representations of statistical concepts that are dynamically linked and offer immediate feedback to their

responses to learner activities. Research at the Open University has looked at how computer-based learning materials can be effectively designed and used to help psychology students come to understand statistical concepts that they find difficult (e.g., Morris, 1998). As a development of this research work, a qualitative evaluation study of *ActivStats* was carried out to see if particular learner activities in this multimedia resource might affect students' understanding of correlation.

The findings of the investigation suggest that the activity that involved students constructing a series of scatter plots was beneficial for students' learning. On this activity, students took turns and commented on each others' actions. In addition, the activity allowed students to manipulate data points on a scatter plot and see the corresponding change in the value of a correlation. During the activity, students also had the opportunity to predict that a particular pattern on a scatter plot would represent a particular value of a correlation coefficient. Moreover, by dragging data points on a scatter plot, students could see the effect that outliers have on a correlation.

The use of multiple representations that are computationally linked in a computer-based learning environment could have important implications in terms of students' learning. It is interesting to note that in Fox and Fowler's (1996) teaching experience students tend blindly to accept the calculated value of the correlation coefficient without considering graphical evidence. Bivariate data should be plotted because isolated points or outliers and clusters of scores can have a powerful influence on the correlation coefficient and would therefore misrepresent the underlying relationship between variables (see Hinton, 1995). In addition, bivariate data can show a curvilinear relationship, but Pearson correlation coefficient, for example, will give a value of zero for the data, indicating no relationship.

The findings of the study suggest that students found it useful to adjust a scale that was linked to a scatter plot, which meant that they could see how a correlation represented data on a graph. However, this activity only involved students moving the sliding scale and they could not directly change the scatter plot to see a corresponding change on the scale.

The findings of the interviews suggest that students thought that the program was useful for learning about correlation. In particular, students thought that the use of scatter plots in *ActivStats* aided their understanding of correlation. For example, when asked about the constructing scatter plots activity, Mark commented, 'again you know the visual impact was good.' During the interviews the students were asked questions to find out about their understanding of the strength of correlations. With the exception of the third pair of students, all of the pairs of students correctly agreed that the correlation -0.8 was stronger than the correlation 0.7. All three pairs of students correctly arranged the set of correlations from the strongest to weakest relationship.

This small-scale study of *ActivStats* provided valuable qualitative data on the process of learning while students worked collaboratively on learning activities for correlation. However, from this qualitative study it is not possible to determine if the learner activities in *ActivStats* significantly contributed to students' understanding of correlation, or addressed particular learner misconceptions because pre- and-post tests were not employed. However, by using an evaluation framework (Jones et al., 1996) that was used in the summative evaluation of *Link*, further empirical work is being carried out with *ActivStats*.

Acknowledgements

References

ActivStats. A Multimedia Statistics Resource (1997), Paul Velleman. Cornell University. Developed by Data Description, Inc., and Addison Wesley Interactive.

Cumming, G. and Thomason, N. (1995), 'Designing software for cognitive change: StatPlay and understanding statistics', in Tinsley, J. D. and van Weerts, T. J. (eds.), *World Conference on Computers in Education VI. WCCE 95, Liberating the Learner*, London: Chapman and Hall, 753–65.

Fox, W. P. and Fowler, C. W. (1996), 'Aiding undergraduate understanding of covariance and correlation', *PRIMUS*, 6 (3), 235–43.

Garfield, J. and Ahlgren, A. (1988), 'Difficulties in learning basic concepts in probability and statistics: implications for research', *Journal for Research in Mathematics Education*, 19 (1), 44–63.

Hinton, P. R. (1995), *Statistics Explained: A Guide for Social Science Students*, London: Routledge.

Issroff, K. (1993), 'Methodology for research in computer-supported cooperative learning', *CALRG Technical Report 132*, Institute of Educational Technology, Milton Keynes: Open University.

Jones, A., Scanlon, E., Tosunoglu, C., Ross, S., Butcher, P., Murphy, P. and Greenberg, J. (1996), 'Evaluating CAL at the Open University: fifteen years on', *Computers and Education*, 26 (1–3), 5–15.

Milheim, W. D. (1995–6), 'Interactivity and computer-based instruction', *Journal of Educational Technology Systems*, 24 (3), 225–33.

Mevarech, Z. R. (1983), 'A deep structure model of students' statistical misconceptions', *Educational Studies in Mathematics*, 14, 415–29.

Morris, E. J. (1997), 'An investigation of students' conceptions and procedural skills in the statistical topic correlation', *CITE Report, No. 230*, Centre for Information Technology in Education, Institute of Educational Technology, Milton Keynes: Open University.

Morris, E. J. (1998), '*Link*: the principled design of a computer-assisted learning program for correlation', in Pereira-Mendoza, L., Seu Kea, L., Wee Kee, T. and Wong, W. (eds.), *Statistical Education – Expanding the Network. Proceedings of the Fifth International Conference on Teaching Statistics. Vol. 2*, Singapore, 21–6 June 1998, International Statistical Institute, 1033–9.

Morris, E. J. (forthcoming), 'The design and evaluation of *Link*: a computer-based learning system for correlation', submitted to *British Journal of Educational Technology*.

Shaughnessy, J. M. (1992), 'Research in probability and statistics: reflections and directions', in Grouws, D. A. (ed.), *Handbook of Research on Mathematics Teaching and Learning. A project of the National Council of Teachers of Mathematics*, New York: Macmillan, 465–94.

Shuell, T. J. (1992), 'Designing instructional computing systems for meaningful learning', in Jones, M. and Winne, P. H. (eds.), *Adaptive Learning Environments*, NATO ASI Series, Vol. F 85, Berlin/Heidelberg: Springer-Verlag, 19–53.