Open Research Online

The Open University's repository of research publications
and other research outputs

The design and evaluation of Link: a CAL system designed to address psychology students' misconceptions about correlation

Thesis

How to cite:

Morris, Erica Jane (1999). The design and evaluation of Link: a CAL system designed to address psychology students' misconceptions about correlation. PhD thesis The Open University.

For guidance on citations see FAQs.

© 1999 The Author

Version: Version of Record

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's data policy on reuse of materials please consult the policies page.

oro.open.ac.uk

UNRESTRICTED

The design and evaluation of *Link*: a CAL system designed to address psychology students' misconceptions about correlation

Submitted for the degree of Doctor of Philosophy

in Educational Technology

January 1999

Erica Jane Morris B.A. (Hons.), M.Sc.

Institute of Educational Technology

The Open University, Milton Keynes, UK

AUTHOR'S No: M7205863

DATE OF SUBMISSION: 28 JANUARY 1999

DATE OF AWARD: 16 AUGUST 1999

Abstract

This thesis describes the design, development and evaluation of *Link*, a computer-assisted learning program for correlation, which is targeted at psychology students in higher education. Computer technology is being increasingly used on statistics courses, suggesting that computer-assisted learning programs on statistical concepts will be increasingly used by students in higher education.

To inform the design of *Link*, an empirical study was conducted to investigate students' difficulties with correlation. It was found that psychology students held misconceptions relating to negative correlations, the strength of correlations and that they infer causality. The design of *Link* was also informed by research-based principles of learning, research and developments in computer-assisted learning and a review of computer-assisted learning programs that cover correlation. A formative evaluation study involving eighteen psychology students found that having used the program, students' general understanding of correlation was significantly improved.

Unlike previously existing computer-assisted learning programs that were reviewed, *Link* makes use of data from two authentic studies in psychology. In addition, *Link* provides learner activities specifically designed to address students' misconceptions about correlation. A summative evaluation study of *Link* involving fifty psychology students was undertaken to assess the effect on students' understanding of correlation. The findings of this evaluation provided further qualitative data on students' misconceptions. Moreover, it was found that the use of *Link* significantly contributed to students' general understanding of correlation.

Contents

Chapter 1	**Introduction**		*1*
	1.1	Background to the thesis	*1*
	1.2	Aims of the thesis	*4*
	1.3	Overview of the thesis	*5*
Chapter 2	**Students' difficulties with statistics**		*10*
	2.1	Introduction	*10*
	2.2	Factors that contribute to students' difficulties with statistics	*10*
	2.3	Students' misconceptions	*13*
	2.4	Statistical misconceptions	*15*
	2.5	Correlation	*24*
	2.6	The assessment of covariation	*33*
	2.7	Summary	*41*
Chapter 3	**An investigation of students' conceptions and skills pertaining to the statistical topic of correlation**		*44*
	3.1	Introduction	*44*
	3.2	Methodology for the investigation	*44*
	3.3	Method	*48*
		3.3.1 Participants	*48*
		3.3.2 Materials	*48*
		3.3.3 Procedure	*51*
	3.4	Findings	*52*
		3.4.1 Statistics as difficult	*52*
		3.4.2 Task booklet	*53*
		3.4.3 Students' answers to the task booklet	*56*
		3.4.4 The piloting of tasks	*73*
	3.5	Discussion	*75*
		3.5.1 Statistics as difficult	*75*
		3.5.2 Tasks to assess students' understanding of correlation	*75*
		3.5.3 Students' difficulties with statistical procedures	*76*
		3.5.4 Students' misconceptions about correlation	*77*
	3.6	Summary	*81*

Chapter 4 **Computer-assisted learning for statistics** *84*
4.1 Introduction *84*
4.2 Learning statistics constructively *84*
4.3 *Stat Lady* *86*
4.4 *StatPlay* *90*
4.5 Computer-assisted learning programs for correlation *94*
4.5.1 *ConStatS* *95*
4.5.2 *Introduction to Research Design and Statistics* *98*
4.5.3 *Statistics for the Terrified* *102*
4.5.4 *Statistics Tutor: Tutorial and Computational Software for the Behavioral Sciences* *106*
4.5.5 *STEPS* *108*
4.5.6 *Understanding Statistics* *113*
4.6 Discussion *114*
4.7 Summary *118*

Chapter 5 **The design and development of *Link*** *120*
5.1 Introduction *120*
5.2 The design of *Link* *120*
5.3 *Link* *126*
5.4 The development of *Link* *132*
5.5 Program implementation *133*
5.6 Using *Link* *137*
5.7 Summary *141*

Chapter 6 **Methodology for the evaluation of computer-assisted learning programs** *142*
6.1 Introduction *142*
6.2 The evaluation of computer-assisted learning programs *142*
6.3 A framework for the evaluation of *Link* *148*
6.3.1 Interactions *149*
6.3.2 Outcomes *150*
6.4 Formative evaluation *150*
6.5 Summative evaluation *153*
6.6 Summary *156*

Chapter 7 Formative evaluation: phase one *157*
7.1 Introduction *157*
7.2 Method *157*
7.2.1 Design *157*
7.2.2 Participants *157*
7.2.3 Materials *158*
7.2.4 Procedure *160*
7.3 Findings *160*
7.3.1 The tasks *160*
7.3.2 Evaluation questionnaire *168*
7.3.3 Additional observations *169*
7.4 Modifications to the first prototype *170*
7.4.1 General modifications to the prototype *170*
7.4.2 Modifications to activity 1 *172*
7.4.3 Modifications to activity 2 *173*
7.4.4 Modifications to activity 3 *173*
7.5 Summary *174*

Chapter 8 Formative evaluation: phase two *175*
8.1 Introduction *175*
8.2 Method *176*
8.2.1 Design *176*
8.2.2 Participants *176*
8.2.3 Materials *177*
8.2.4 Procedure *179*
8.3 Findings *180*
8.3.1 Interactions: the program *180*
8.3.2 Outcomes: tests in correlation *187*
8.3.3 Outcomes: questions in correlation *195*
8.3.4 Interactions: case studies *203*
8.3.5 Outcomes: participants' opinions of the program *218*
8.4 Discussion *220*
8.5 Expert evaluation of the program *226*
8.6 Summary *228*

Chapter 9 A summative evaluation study of *Link* *229*
9.1 Introduction *229*
9.2 Final design of *Link* *229*
9.2.1 Activity 1 *234*
9.2.2 Activity 2 *237*
9.2.3 Activity 3 *239*
9.3 Introduction to summative evaluation *241*
9.4 Method *242*
9.4.1 Design *242*
9.4.2 Participants *242*
9.4.3 Materials *244*
9.4.4 Procedure *248*
9.4.5 Procedure for case studies *249*
9.5 Findings *249*
9.5.1 Students' interactions *250*
9.5.2 Outcomes *256*
9.5.3 Students' misconceptions *259*
9.5.4 Case studies *271*
9.6 Discussion *278*
9.7 Summary *283*

Chapter 10 Conclusions *285*
10.1 Introduction *285*
10.2 Achievements *285*
10.3 Implications for research *290*
10.4 Implications for computer-assisted learning for statistics *292*
10.5 Implications for education *293*
10.6 Limitations of the research *294*
10.7 Further improvements to *Link* *295*
10.8 Further research *296*
10.9 Summary *298*

References *299*

Appendices

Appendix A Participant instructions for the investigation *311*

Appendix B Questions for task booklet *312*

Appendix C A sample of Lingo scripts *329*

Appendix D Evaluation questionnaire *335*

Appendix E Tests in correlation used for the formative evaluation *337*

Appendix F Final design of *Link* *352*

Appendix G A sample user log *375*

Appendix H Applications *377*

Figures

Figure 2.1 Scatter plots of data sets from Anscombe (1973)
Figure 3.1 Frequency distribution of students' scores on the task booklet
Figure 4.1 The *Sampling Playground* in *StatPlay*
Figure 4 2 *Describing Bivariate Data* in *ConStatS*
Figure 4.3 *What is correlation?*
Figure 4.4 A learner activity in *What is correlation?*
Figure 4.5 A student is asked to match the description of a relationship with a scatter plot
Figure 4.6 Feedback to a question in *What is correlation?*
Figure 4.7 The correlation coefficient is introduced in *Statistics for the Terrified*
Figure 4.8 An exercise with 94 seconds remaining
Figure 4.9 Correlation is summarised in *Statistics for the Terrified*
Figure 4.10 An exercise that involves moving data points on a scatter plot
Figure 4.11 An exercise from the *Correlation Tutorial*
Figure 4.12 A scatter plot can be generated in *Predicting Dyslexia?*
Figure 4.13 A correlation coefficient is displayed for a pair of variables
Figure 4.14 General examples of correlation
Figure 4.15 An exercise from *Understanding Statistics*
Figure 5.1 The introductory screen of the first prototype
Figure 5.2 Introductory screen for the TV violence study
Figure 5.3 A screen that provides details of the TV violence study
Figure 5.4 Schema that describes the organisation of the first prototype
Figure 5.5 Activity 2
Figure 5.6 Activity 1. Options are highlighted in accord with relevant audio feedback
Figure 5.7 Activity 3
Figure 5.8 An additional screen for activity 3
Figure 7.1 Activity 1
Figure 7.2 Activity 2. The correlation -0.65 has been selected
Figure 7.3 Activity 3
Figure 7.4 Additional screen for activity 3
Figure 7.5 Activity 2. For the second prototype the buttons at the bottom of the screen were changed
Figure 7.6 Activity 1 in the second prototype
Figure 7.7 Activity 3 in the second prototype

Figure 8.1 Frequency distribution of participants' pre test scores (excluding filler questions)
Figure 8.2 Frequency distribution of participants' post test scores (excluding filler questions)
Figure 8.3 Frequency distribution of participants' pre test whole question scores (excluding filler questions)
Figure 8.4 Frequency distribution of participants' post test whole question scores (excluding filler questions)
Figure 8.5 McNemar test: question 14 (cases 12)
Figure 8.6 McNemar test: question 4 (cases 18)
Figure 9.1 Schema for *Link*
Figure 9.2 Health events study
Figure 9.3 Health events study. A pair of variables has been selected
Figure 9.4 Activity 1
Figure 9.5 Health events study. Feedback to activity 1
Figure 9.6 Infant engagement study. Feedback to activity 1
Figure 9.7 Infant engagement study. Activity 2
Figure 9.8 Health events study. Activity 2
Figure 9.9 Health events study. Feedback to activity 3
Figure 9.10 Health events study. Activity 3
Figure 9.11 Infant engagement study. Activity 1
Figure 9.12 Infant engagement study. Activity 3
Figure 9.13 Pre and post test means of the three groups

Figures from appendix B

Figure 1 Scatter plot of study time by exam. performance
Figure 2 Scatter plot of reading scores by arithmetic scores
Figure 3 Correlations between the amount of television violence viewed at age 9 and peer-rated aggression at age 9 (184 boys and 175 girls)
Figure 4 Correlations between the amount of television violence viewed at ages 9 and 19 and peer-rated aggression at ages 9 and 19 (184 boys and 175 girls)

Tables

Table 1.1 Timetable of research

Table 3.1 Answer categories for questions: per cent of student responses for particular categories

Table 6.1 Framework for evaluation. (Adapted from Jones et al, 1996, p. 12)

Table 6.2 Framework for the second phase of the formative evaluation

Table 6.3 Framework for the summative evaluation

Table 8.1 Descriptive statistics for overall scores of pre and post tests

Table 8.2 Descriptive statistics for overall scores of pre and post tests (filler questions excluded)

Table 8.3 Descriptive statistics for whole question scores of pre and post tests

Table 8.4 Descriptive statistics for whole question scores of pre and post tests (filler questions excluded)

Table 9.1 Participant descriptive statistics for the three groups

Table 9.2 Participant responses to activity 2. (Health events study)

Table 9.3 Participant responses to activity 3. (Health events study)

Table 9.4 Participant responses to activity 2. (Infant engagement study)

Table 9.5 Participant responses to activity 3. (Infant engagement study)

Table 9.6 Descriptive statistics for the pre and post tests

Table 9.7 ANCOVA summary table

Table 9.8 Participant responses to questions on the pre test (*Link* group)

Table 9.9 Participant responses to questions on the pre test (Instructional control group)

Tables from appendix B

Table 1 Students' study time and examination marks

Table 2 Children's arithmetic and reading scores

Table 3 Subjects' memory test scores and spelling test scores

Acknowledgements

I would like to give a big thank you to my supervisors, Eileen Scanlon and Ann Jones. They have provided sound advice, discussion, encouragement and support throughout my PhD work and to them I am extremely grateful.

Many thanks must go to Jane Barnard for being such a good friend and office mate, and for her kindness, help, advice and excellent humour. I would like to thank Pat Fung, Cathy Lewin, Martin Oliver and Martin Le Voi for reading earlier versions of my thesis and for providing valuable comments.

I would like to thank members of the Institute of Educational Technology who have helped make it such a friendly and supportive environment for learning as a postgraduate. In particular, I would like to thank Gill Kirkup for helpful discussions on my yearly progress, and Pauline Adams, Hansa Solanki and Di Mason for their help.

Anne Hawkins at the Centre for Statistical Education was most helpful during my research for supplying references to follow up and her help is much appreciated. A thank you to Geoff Cumming for sending me his research papers and for letting me use a screen shot of *StatPlay* in my thesis. I am also very grateful to John Richardson for taking an interest in my research and for providing constructive criticism and thoughts. A big thank you must go to Richard Joiner: for his encouragement, for letting me discuss my ideas and research problems and for helping me with those bits of statistics. Many thanks must go again to Martin Le Voi who has provided advice and support throughout my thesis work.

I am very grateful to Karen Vines who provided statistical advice. I would like to thank Chris Jones and Andrew Bertie who also provided statistical expertise and advice particularly during the earlier stages of my research.

My research involved visiting universities so that I could involve students in my studies and I am most grateful to Eddie Shoesmith, Tony Ward and Peter Naish for their help in letting me contact students and for supporting my work in this way. A big, big thank you to all those students who were involved in my studies. A thank you to Rupa Kundu and Anna Campbell who provided friendship, conversation and tea while I sat in a psychology lab. running my final study.

I am very grateful to Gillian Cohen and John Oates who were a great help in my search for real data and for letting me use their studies and data sets for my program. For programming advice I would like to thank Cathy Lewin and Matthew Stratfold.

I would like to thank the following for their friendship, humour, support and encouragement: Cathy Lewin, Stylianos Hatzipanagos, Martin Oliver, Iraklis Paraskakis, Rick Holliman and Ed Hall. A big thank you must also go to Tina Bowen, Fenella Potterton and Rachel Evans for their support and friendship and for much needed long telephone conversations. An affectionate thank you to Trevor Collins for his help and support when my viva was in sight.

I would like to give many thanks to my parents, Jacqui Morris and Brian Morris who have shown me how to keep my feet on the ground, and to a special daughter, Tallulah Morris who has always helped me keep things in perspective.

Chapter 1

Introduction

1.1 Background to the thesis

The research in this thesis concerns the development and evaluation of a computer-assisted learning program, called *Link*, for the statistical topic of correlation, which is targeted at students taking undergraduate degree programmes in psychology. Table 1.1 summarises the research activities that contributed to the thesis (p. 9).

A psychology degree in higher education requires that students take courses in descriptive and inferential statistics, but research reviewed in this thesis indicates that students can find statistics difficult. The focus of this thesis is the topic of correlation, and the kinds of difficulties and confusions that students have concerning this area. Correlation is a fundamental statistical topic for psychology students that is typically covered in introductory statistics courses. Students need to interpret correlational data from psychology studies in the research literature and carry out projects that involve the collection, description, analysis and interpretation of data. For these projects a hypothesis might be proposed to investigate whether a relationship exists between variables.

The statistics curriculum is changing because of the increasing use of computers in higher education (Hawkins, Jolliffe & Glickman, 1992). For example, students do not necessarily have to learn how to calculate a statistic, but can use data-analysis programs, such as *SPSS* (*SPSS*, 1989 - 1995) to handle a large data set and generate appropriate statistics. Such changes have begun to affect the teaching and learning of statistics and the focus can now be on understanding statistics as opposed to computation (Hawkins et al, 1992). Pierce and Roberts (1998) have described a statistics service course that was changed in response to advances in computer technology. Prior to the 1990s, the focus of many statistics courses was the application of statistical techniques and so using

formulas, carrying out calculations and learning short cuts to calculate statistics took up much of the teaching (Pierce & Roberts, 1998). However, access to data-analysis software has meant that data sets can be analysed in relatively short amounts of time and this has allowed teachers and students to concentrate on understanding the principles and concepts of statistics (Pierce & Roberts, 1998). The statistics course that Pierce and Roberts (1998) have outlined has included: lectures, where relevant concepts are introduced; computer laboratory classes, where students use data-analysis packages; problem classes, which use computer output from the laboratory classes; and discussion classes, where students discuss statistical issues raised by newspaper or magazine articles.

Computer technology is being used increasingly in higher education. This increase is likely to apply to the use of both data-analysis software and computer-assisted learning programs by psychology students on their statistics courses. Taylor's (1980) framework for classifying educational computing is relevant in this context: the computer can function as a tutor, tool or tutee. One mode for using computing in education is for the computer to impart subject matter. In this tutor mode, the computer presents subject material, the student responds, (for example, to questions) the program evaluates the response and determines what to subsequently present (Taylor, 1980). There are, inevitably variations from this scheme. The computer can also be used as a tool in education where it "need only have some useful capability programmed into it such as statistical analysis" (ibid., p. 3). The third mode suggested by Taylor is where the computer is used as a tutee, where the student instructs the computer by, for example, writing programs. There are programming languages that might help students to learn about statistics because they can write code to derive statistical concepts, such as the mean, but it has been argued that this is not a very efficient way of using students' time (Hawkins et al, 1992). The research described in this thesis looks at computer-assisted learning programs for statistics and so the use of the computer as a tool or a tutee are beyond the scope of the research. With regard to Taylor's (1980) classification, a

computer-assisted learning program is defined in this thesis as a piece of educational software that was designed to present, impart and/or review a particular chunk of subject matter, principles or concepts. With the advances of computer technology, such programs might incorporate multimedia elements and allow the student to choose which parts of the program they wish to work through. In statistics education, a computer-assisted learning program must be distinguished from a data-analysis program that can be used to organise and analyse data. More recently, Biehler (1994) has noted that for introductory statistics education, several kinds of software are relevant and in use and these include custom designed programs for a specific educational goal and statistical systems for data analysis. Taylor's (1980) succinct framework still holds. There are integrated applications, such as *ActivStats* (*ActivStats*, 1997) that provide both a resource designed to teach statistical concepts and techniques and data-analysis software. The Data Desk software, which allows the student to analyse data sets, is linked to *ActivStats*.

The field of computer-assisted learning is usually viewed as being concerned with the development of intelligently designed educational software, which is produced to solve essentially educational goals by computational means (du Boulay, 1998). In this thesis it is argued that the design of computer-assisted learning programs should be informed by research-based principles of learning. Chapter 4 of this thesis reviews work that has emphasised that learning is cumulative and is a constructive process (e.g., Shuell, 1992) and that the acquisition of concepts is facilitated if they are anchored to realistic contexts that are meaningful to the learner (e.g., Bransford, Sherwood, Hasselbring, Kinzer & Williams, 1990). This means that the learner's prior knowledge must be addressed in the design of computer-assisted learning materials and that material to be learnt should be presented in the context of interesting, real world examples. Empirical work on students' understanding of a variety of subject matter areas has enabled students' misconceptions to be identified (chapter 2). Ohlsson (1991) has argued that error analysis, in which students' common errors or misconceptions are identified through research, remains an essential component of instructional design in general, and

of system design in particular.

Drawing on research perspectives, it is also contended in this thesis that the design of a computer-assisted learning program should harness the instructional capabilities of computer technology. As Shuell (1992) has suggested: it is important to consider the capabilities of computers that provide advantages with regard to their use for instructional purposes.

In summary, it is likely that computer-assisted learning programs are going to be used increasingly in higher education and that psychology students are likely to use them on their statistics courses. Psychology students might find statistics a difficult subject to study and a computer-assisted learning program could provide an additional form of instruction to help students acquire statistical concepts. It is argued in this thesis that the design of computer-assisted learning programs should be informed by research-based principles of learning, empirical studies that have identified students' misconceptions, research and developments in the field of computer-assisted learning and recommendations for the effective design of such programs. Furthermore, the design and development of a computer-assisted learning program should involve the formative evaluation of the program with target users. In this context, the focus of the research described in this thesis is the topic of correlation and the development of a computer-assisted learning program that was designed to address students' difficulties with this area.

1.2 Aims of the thesis

The research described in this thesis had four primary aims as follows:

- To investigate students' difficulties and confusions pertaining to the area of correlation.
- To design a computer-assisted learning program for the area of correlation that was based on research in student learning, and research and developments in the field of

computer-assisted learning.

- To see whether the program *Link* contributed to students' general understanding of correlation.
- To investigate whether learner activities in *Link* addressed students' confusions that relate to their understanding of correlation.

1.3 Overview of the thesis

The first empirical study that was conducted for the thesis research was designed to identify the confusions and difficulties that students might have with the statistical topic of correlation. The findings of this investigation, which formed the basis of the design and development of the program called *Link,* indicated that some psychology students hold misconceptions relating to causality, negative correlations and the strength of correlations. Accordingly, the program was designed to address these misconceptions. More specifically, the design of *Link* was based on:

- Relevant research (e.g., empirical studies that concern students' misconceptions about correlation).
- Research and developments in computer-assisted learning for statistics.
- A formative evaluation of the program.
- An expert evaluation of the program.

Chapter 2 reviews the research that concerns students' difficulties with statistics. The reasons why students find statistics a difficult subject to master are outlined. These include mathematical skills and affective factors. A great deal of research has looked into students' misconceptions in statistics and a variety of empirical disciplines and perspectives have contributed to this field of research.

Chapter 2 also provides a rationale for why the topic of correlation was chosen as the focus of empirical research described in this thesis. Correlation is described with

regard to it being part of the typical curriculum for psychology students. A critical review of the research that relates to students' difficulties concerning correlation is also provided in chapter 2.

Chapter 3 provides an account of an investigation of students' conceptions and skills in the topic of correlation. This study had three primary objectives. Firstly, it was designed to investigate students' experience in studying statistics in terms of, for example, their interest and difficulty. Secondly, an objective of the study was to investigate the kinds of confusions and difficulties that students experience with correlation. Thirdly, the study was set up to pilot questions that were devised to assess students' understanding of this topic.

Chapter 4 critically examines research and developments in the field of computer-assisted learning for statistics. Two research programmes are particularly pertinent here: *Stat Lady* (Shute, Gawlick-Grendell, Young & Burnham, 1996) and *StatPlay* (Cumming & Thomason, 1998). The design of *Stat Lady* has been based on mastery learning and empirical work has focused on the evaluation of the system's probability module. *Stat Play* has been designed to address students' naive statistics by the use of simulations, demonstrations and dynamically linked representations of statistics concepts. Through an examination of these two research programmes important issues that relate to the design and evaluation of computer-assisted learning for statistics are discussed.

Chapter 4 also describes a review of computer-assisted learning programs for statistics that was conducted, which looked at how different programs present the topic of correlation. This review provides recommendations for the design of a computer-assisted learning program for correlation.

Chapter 5 describes the design and development of *Link*. The design of the first prototype of *Link* was informed by relevant research outlined in chapter 2 and chapter 3. The design of *Link* was also based on an examination of research and developments in

the field of computer-assisted learning for the topic of correlation (chapter 4). This provided suggestions for possible learner activities that could be used in a computer-assisted learning program.

In Chapter 5 the approach taken in the development of *Link* and the authoring environment, Macromedia Director, used in this process are described. A first prototype was originally developed to be used in a formative evaluation study of the program.

The effective design and development of a computer-assisted learning program should involve an evaluation of the application. Chapter 6 looks at pertinent literature on the evaluation of computer-assisted learning in higher education. This literature provides recommendations concerning the methods that should be used in an evaluation study and advocates that students, or the anticipated users of the program, must be involved in the evaluation process. A methodology for the evaluation of a computer-assisted learning program is outlined that was employed in the development and evaluation of *Link*.

Chapter 7 describes the first phase of a formative evaluation study of *Link*. The focus of this evaluation was an assessment of the usability of the program. This study provided qualitative data concerning, for example, students' difficulties while they used the program, which was used to delineate modifications to the first prototype. Based on the findings of this first study, a second improved prototype of *Link* was produced.

Chapter 8 reports on the second phase of the formative evaluation of *Link*. The objectives of this study were to:

- Investigate whether *Link* contributed to students' general understanding of correlation.
- Investigate whether *Link* affected students' misconceptions about correlation.
- Provide a formative evaluation of the program's learner activities and presentation of topic material.

- Pilot tests that were designed to provide an assessment of students' understanding of correlation. These tests could then be used in the summative evaluation study.

Chapter 8 describes the findings of the formative study which indicated that the program significantly contributed to students' general understanding of correlation. As part of the evaluation process, expert evaluation of *Link* was also conducted, providing valuable qualitative data concerning how the program could be improved. The findings of the second phase of the formative study and the expert evaluation were used to inform further modifications to *Link.*

Chapter 9 describes the final implementation of *Link* and a summative evaluation of this program. The findings of this evaluation indicated that *Link* significantly contributed to students' general understanding of correlation.

Chapter 10 describes the achievements of the research described in the thesis and the implications that this work has for research, education and for computer-assisted learning programs. Here, limitations of the research are outlined, and improvements to the design of *Link* and further research possibilities that could be undertaken are considered.

Table 1.1 Timetable of research

Year	Research activity	Outcome	Chapter
Year 1	Investigation of students' conceptions and skills in the topic of correlation.	Questions were piloted to test students' understanding of correlation. Students' misconceptions in correlation were identified.	Chapter 3.
Year 2	Evaluation of computer-assisted learning programs.	Learner activities that could be used in a program for correlation were identified.	Chapter 4.
	Development of first prototype.	Prototype of *Link*.	Chapter 5.
	Formative evaluation study: phase one.	Problems that related to the usability of *Link* were identified.	Chapter 7.
	Development of second prototype.	Second prototype of *Link*.	Chapter 7.
	Formative evaluation study: phase two.	After using *Link* students' understanding of correlation was significantly improved. Data that indicated which aspects of the learner activities must be modified.	Chapter 8.
	Expert evaluation.	Recommendations for the improved design of *Link*.	Chapter 8.
Year 3	Development of final program.	*Link*.	Chapter 9.
	Summative evaluation study.	Students' general understanding of correlation improved Further data on students' misconceptions	Chapter 9.

Chapter 2

Students' difficulties with statistics

2.1 Introduction

This chapter examines literature that helps to explain why some students taking undergraduate degree programmes in psychology might find statistics difficult. It is important to note that much of the literature that is relevant to why students have difficulties with learning statistical topics does not necessarily concern psychology students. Research in mathematics and science learning has shown that students have misconceptions relating to a variety of subject areas. In an attempt to ascertain why some psychology students might find statistics difficult, research is reviewed that has indicated that people tend to hold statistical misconceptions in particular topic areas.

The focus of the research described in this thesis concerns the statistical topic of correlation and this chapter looks at this topic as a part of the statistical curriculum for those students taking psychology at undergraduate level. Accordingly, research that has looked at novice and expert detection and assessment of covariation (e.g., Well, Boyce, Morris, Shinjo & Chumbley, 1988) is also reviewed.

2.2 Factors that contribute to students' difficulties with statistics

As in other areas of science and mathematics, students who study statistics need to acquire mathematical skills and subject matter concepts. It is therefore likely that the difficulties students encounter in learning statistics are not particularly different from the problems encountered by students who study an area of science or mathematics. However, statistics is relevant to a variety of disciplines (Hawkins et al, 1992), which means that students who are studying psychology, geography, economics, or a biological science are required to learn to use statistics to describe, analyse and interpret data. This is likely to create problems in students' learning because, for example, statistics will not

necessarily be a subject that students have chosen to study and this might mean that students lack the motivation to study statistics. In addition, the statistics curriculum might not use problems that are directly relevant to the students' chosen discipline.

Garfield and Ahlgren (1988) have highlighted the interrelated factors that are likely to contribute to students' difficulties in mastering statistical concepts and techniques: affective factors, such as a lack of motivation to attend to statistical topics, inadequacies in prerequisite mathematical skills, and statistical misconceptions. Comprehensive reviews of research that concern students' difficulties in statistics (Garfield & Ahlgren, 1988; Shaughnessy, 1992) have not, however, referred to particular groups of students who must study statistics as part of an undergraduate degree programme in a subject discipline, such as psychology or economics. Rather, such reviews have highlighted the finding that people in general (including students) tend to find statistics difficult to understand. It is an empirical question whether psychology students lack particular prerequisite skills for statistical techniques or suffer a host of statistical misconceptions that impede the acquisition of statistical concepts. This question was addressed in the empirical study that is described in the following chapter. Research concerning students' difficulties with statistics has tended to focus on statistical misconceptions and not much attention has been paid to, for example, affective factors that might contribute to students' success at learning a statistical topic. Related to this is a lack of research on whether students subjectively find statistics difficult or whether they do not find it an interesting topic to study. Researchers have tended to report on educators' experience that students find statistics difficult rather than looking at empirical findings:

> "The experience of most college faculty members in education and the social sciences is that a large proportion of university students in introductory statistics courses do not understand many of the concepts they are studying" (Garfield & Ahlgren, 1988, p. 46).

Garfield and Ahlgren (1988) have pointed out that studies in the research literature supported this view, but such studies concern students' lack of understanding of statistical topics rather than whether students themselves find statistics difficult or uninteresting.

It is unclear why research has not tended to focus on students' inadequacies in prerequisite mathematical skills as a possible contributing factor to students' mastery of statistical topics. Specifically, research studies have not usually considered how students' mathematical skills might affect their success in understanding statistical concepts. However, the statistics curriculum is changing with the increasing use of computers in higher education (Hawkins et al, 1992), which is likely to mean that more students will use data-analysis applications to compute statistics. For example, psychology students will make increasing use of *SPSS* (*SPSS*, 1989 - 1995) or *StatView* (Abacus Concepts Inc., 1992 - 1993) to analyse data for statistics and will not necessarily learn how to compute statistics by hand by following a predefined procedure in a textbook (e.g., Minium, 1978, p. 338). This might mean that prerequisite mathematical skills are not so important in the learning of statistics because students will not have to use formulas to calculate statistics, but can use data-analysis software for this purpose.

2.3 Students' misconceptions

As will be considered in chapter 4 of this thesis, there is a consensus of opinion that learning is cumulative (De Corte, 1995), which means that students construct new knowledge on the basis of their prior knowledge. It is argued in chapter 4 that the development of computer-assisted learning programs should be principled: students' prior knowledge in the form of misconceptions must be taken into account in the design of such programs.

There has been a wealth of research that has identified students' misconceptions in a variety of subject areas in science and mathematics (Smith, diSessa & Roschelle, 1993). For example, Songer and Mintzes (1994) have explored and documented students' conceptual difficulties concerning the processes of cellular respiration by using concept maps and clinical interviews. Smith and his associates have pointed out that this kind of research has been valuable because it has produced detailed characterisations of students' understandings of subject areas and this has represented an advance from previous approaches to learning that simply divided student responses into correct or incorrect categories.

Research has shown that students have conceptions that explain some mathematical and scientific concepts, but such conceptions are different from the formal concepts presented in instruction (Smith et al, 1993). However, as Smith et al have pointed out, education in mathematics and science needs to take these students' conceptions seriously because they regularly differ from formal concepts and can guide students' reasoning. Although misconceptions in a variety of subject areas have been found to be persistent and resistant to instruction, not all misconceptions are stable and resistant to change (Smith et al, 1993).

Nevertheless, there is a consensus of opinion that in the design of instruction, students' prior knowledge in the form of common errors or misconceptions must be

addressed (Driver, 1988; Hennessy, Twigger, Driver, O'Shea, O'Malley, Byard, Draper, Hartley, Mohamed & Scanlon, 1995; Laurillard, 1993; Shuell, 1992). Much of the focus of research in mathematics and science learning has been to document misconceptions in different domains and less emphasis has been given to describe the kinds of instruction that successfully promotes learning (Smith et al, 1993). This situation is evident when research relating to statistical misconceptions is discussed below: with the exception of Mevarech (1983), there has been a lack of research that has looked at what learning conditions address students' statistical misconceptions and promote the acquisition of concepts.

In the case of science, however, Driver (1988) has described a project that was undertaken to devise, trial and evaluate constructivist teaching sequences for particular topic areas. These sequences were designed to take account of students' prior conceptions and to promote conceptual change. Within the teaching sequences developed by the project, a number of different strategies were used to facilitate the construction of new concepts (Driver, 1988). These teaching strategies included the following:

- Broadening the range of application of a conception.

  Here, students' prior conceptions can be used as a resource which can be extended.

- Differentiation of a conception.

  In some topic areas, students' conceptions are global and ill-defined and particular experiences are necessary to help them differentiate their ideas.

- The construction of an alternative conception.

  Students' prior conceptions can be incommensurate with formal conceptions and so problems might arise if these prior conceptions are used to shape new concepts. In such cases, students' prior ideas are acknowledged and discussed and the alternative scientific model is put forward. Students have the opportunity to evaluate the scientific model in relation to their prior conceptions.

  (Adapted from Driver, 1988, pp. 143 - 145).

In the design of teaching sequences, Driver (1988) has pointed out that the choice of a particular strategy has depended on the kind of prior conception that students hold and the learning objectives. Driver (1988) has suggested that strategies, which are designed to promote conceptual change, need to be investigated in the context of particular areas of knowledge.

Misconceptions research has resulted in a variety of terms to describe students' conceptions that are at odds with accepted theory, such as preconceptions and alternative conceptions (Smith et al, 1993). This is discussed below because research into students' understanding of statistical concepts has resulted in a variety of terms to describe students' systematic errors. The term misconception is used in this thesis to refer to a student conception that describes a pattern of errors and in the case of statistics, to refer to a conception that is not consistent with accepted statistical theory.

2.4 Statistical misconceptions

A variety of different areas of enquiry have indicated that lay people, students and even researchers hold statistical misconceptions (Shaughnessy, 1992). Indeed, Cumming and Thomason (1995) have asserted that misconceptions in statistics are "widespread, persistent" and "resistant to conventional education" (p. 754), but these researchers do not cite evidence to support these assertions. Multidisciplinary efforts have shown that people's ideas are typically at odds with statistical theory, and because different fields of

enquiry have uncovered people's statistical confusions, numerous terms have been used to describe statistical misconceptions. One finds: naive statistics and faulty intuitions (Cumming & Thomason, 1995), errors and misconceptions (Mevarech, 1983) and preconceptions and misunderstandings (Shaughnessy, 1992). A distinction can be made between students' prior conceptions, which they bring to the learning of statistics and misconceptions, which they might have acquired in their learning of statistics. Laurillard (1993) has pointed out that students may possess pedagogenic errors that are teacher (or textbook) induced. Indeed, Brewer (1985) has documented inaccuracies and errors in behavioural statistics textbooks that reflect misconceptions of statistical theory and might mislead the behavioural researcher (or student). Brewer (1985) examined best selling introductory textbooks because he thought:

> "that the first exposure to statistics is where a form of 'misconception imprinting' takes place, crucially affecting the researchers' statistical beliefs ... for years to come" (ibid., p. 255).

For example, a statement in one statistics textbook indicated a misconception relating to the alpha levels of 0.01 and 0.05. The implication of the statement was that there are only two alpha levels, but there is no magic in 0.01 or 0.05 because any alpha level not equal to zero or 1 may be selected by a researcher (Brewer, 1985).

The situation is therefore complex: given a particular student it is not clear if, for example, a prior conception remains even after the student has taken courses in statistics or whether the student acquires a particular misconception through instruction. To confuse matters further, students might hold certain normal conceptions and misconceptions in statistics simultaneously. In Shaughnessy's (1992) micro-model of stochastic conceptual development, which comprises four levels of stochastic understanding: non-statistical, naive-statistical, emergent-statistical and pragmatic-statistical, to characterise how students develop, hold and apply models of probability, it is emphasised that the levels are not necessarily linear or mutually exclusive. You do not

have to be naive-statistical before you are pragmatic-statistical and you can function with several of the different levels or categories operative (Shaughnessy, 1992).

Much research has concerned people's statistical misconceptions, but there is much less research that has looked at how certain learning conditions can remedy such confusions. One exception here is the work by Mevarech (1983) that is described later in this section. With regard to research related to students' difficulties with statistics, far more research has concerned probability than other statistical concepts (Garfield & Ahlgren, 1988). An area of enquiry in psychology referred to as judgment under uncertainty has illustrated that misconceptions concerning probability are commonly held (Kahneman & Tversky, 1972; Konold, 1989; Tversky & Kahneman, 1982). Tversky and Kahneman (1982) proposed that people tend to use judgmental heuristics in uncertain situations in which a decision must be made. More specifically, it has been argued that people do not adhere to probability theory when judging the probabilities of uncertain events, but use heuristics that are difficult to eliminate (Kahneman & Tversky, 1972). For example, in tasks that require an assessment of probability, individuals typically employ the representativeness heuristic where a person

> "evaluates the probability of an uncertain event, or a sample, by the degree to which it is: (i) similar in essential properties to its parent population; and (ii) reflects the salient features of the process by which it is generated ... in many situations, an event A is judged more probable than an event B whenever A appears more representative than B" (Kahneman & Tversky, 1972, p. 431).

By relying on this heuristic an individual will therefore consider that a sample will reflect the characteristics of the population from which it is drawn. This is illustrated by the following problem:

"All families of six children in a city were surveyed. In 72 families the exact order of births of boys and girls was GBGBBG. What is your estimate of the number of families surveyed in which the exact order of births was BGBBBB?" (ibid., p. 432).

Kahneman and Tversky (1972) reported that 75 out of 92 participants judged the sequence BGBBBB to be less likely than the sequence GBGBBG although the two birth sequences are equally likely. In other words, the sequence GBGBBG is viewed as representative of the proportion of boys and girls in the population. Misconceptions that concern probability are not only held by novices, but also by trained scientists (Tversky & Kahneman, 1993). Although this type of research has indicated the kinds of misconceptions about probability that people tend to suffer from, it does not provide data concerning the kinds of difficulties that students might typically encounter when they study statistics as part of an undergraduate degree programme in a user-discipline of statistics, such as psychology. There are a variety of statistical topics that psychology students must study in descriptive and inferential statistics. Although less attention has been given to research on misconceptions in these topics areas (Garfield & Ahlgren, 1988) research has suggested that students also have conceptual difficulties in descriptive statistics. Descriptive statistics are used to organise and summarise samples of data, such as measures of central tendency.

Researchers have argued that many students do not have a complete and thorough understanding of the arithmetic mean (Hardiman, Well & Pollatsek, 1984; Pollatsek, Lima & Well, 1981). This argument has been based on findings that indicated that students' knowledge of the mean was limited by "an impoverished computational formula" (Pollatsek et al, 1981, p. 191), which can be used to solve simple mean problems, but not weighted mean problems (ibid.) To assess students' understanding of the mean, Pollatsek et al (1981) interviewed students and also asked them to think aloud while they worked on weighted mean problems. So, the students (most of whom were

undergraduate psychology students) were, for example, asked to solve the following problem:

> "A student attended college A for two semesters and earned a 3.22 GPA [grade point average]. The same student attended college B for four semesters and earned a 3.78 GPA. What is the student's GPA for all his college work?" (ibid., p. 195).

By using such problems, Pollatsek et al (1981) found that many students were unable to correctly weight and combine two means to give a single mean. With regard to the above problem, students unthinkingly apply the computational formula for a single mean and add 3.22 and 3.83 and then divide the outcome by 2. They should calculate 2 multiplied by 3.22 and add this result to 4 multiplied by 3.78, and then divide by six (the total number of semesters). Indeed, of the fifteen students who worked on a grade point average problem, only two students computed the right answer and the others tended to take the unweighted mean of the two GPA's even when the hypothetical student spent twice as much time at college B than at college A. Thus, in a typical interview with a student solving a grade point average problem, the student would add the GPA for college A and B and then divide the result by 2 (Pollatsek et al, 1981). With these findings in mind, Pollatsek et al (1981) suggested that students' knowledge of the mean was limited to a computational formula, which was not sufficient for a complete conceptual understanding of the mean. Moreover, these researchers noted that in many introductory courses in statistics students learn to use formulas in a rote like manner and that non mathematical students tend to focus on the learning of formulas to solve specific statistical problems. The study, however, only involved a sample of seventeen students and just three of these participants had completed about half a semester of statistics before taking part in the study (Pollatsek et al, 1981). The study was also somewhat limited in its scope because it exclusively looked at students' solving weighted mean problems. One of the problems used in the study read as follows:

> "There are ten people in an elevator, four women and six men. The average weight of the women is 120 pounds, and the average weight of the men is 180 pounds. What is the average of the weights of the ten people in the elevator?" (Pollatsek, 1981, p. 195).

Hawkins and her colleagues (1992) have noted that it is not clear why researchers would want students to calculate such a pointless statistic because it expects students to work out a representative statistic for a problem that has two different distributions.

Mevarech (1983) took a rather different approach in looking at students' understanding of descriptive statistics. In one of his studies, students acted as diagnosticians who were presented with a test of statistical problems that had been solved either correctly or incorrectly. As diagnosticians, the students were asked to identify whether the problems had been solved incorrectly, to describe the erroneous steps and to propose appropriate corrections. Mevarech (1983) found that all the students, who had taken courses in descriptive statistics, possessed prerequisite computational knowledge in that they recognised formulas for the simple mean, weighted mean and variance. However, this knowledge was not sufficient to solve a variety of given problems in descriptive statistics because only a few students possessed the necessary conceptual structures to solve the problems in the diagnostic modelling test (Mevarech, 1983). For example, one of the problems on the test was designed to diagnose a misconception that concerns the 'identity number':

"A score of zero (0) was added to a set of 5 scores (52, 68, 74, 86 and 90) with a mean equal to 74. What is the mean of the new set?

(Incorrect) solution: The mean will not be changed because adding zero to the sum does not change the sum. Thus,

$(52 + 68 + 74 + 86 + 90 + 0)/5$

$= (52 + 68 + 74 + 86 + 90)/5$

$= 370/5 = 74$"

(Mevarech, 1983, p. 418).

In the above case, the solution is incorrect because the score 0 should be added to give a sum, and this sum should be divided by 6, the new number of scores in the set. This ensures that the value of the mean is influenced by the value of every score in a particular distribution. However, 30 per cent of the students evidently did not understand this and thought that zero was the identity element or that when it was added to a set of scores it would not change the mean (Mevarech, 1983). In the operation of addition, the identity element is zero, (which means that the sum of zero and any other number is the number itself), and it is proposed that students inappropriately use this knowledge in working out the average of a set of numbers (Mevarech, 1983). Based on an analysis of students' errors, Mevarech (1983) proposed that most of the students who took part in his research did not seem to fully understand the idea of using average numbers and variances.

Mevarech (1983) also carried out a study that was designed to investigate whether a Mastery Learning Strategy helped students to overcome statistical misconceptions. More specifically, this study investigated whether students exposed to instruction with a Mastery Learning Strategy out-performed those students in a control group who had been exposed to a traditional lecture course. Both groups of students covered the following topics on the courses: frequency distributions, measures of central tendency, measures of dispersion and correlations. The students who took part in the experimental group were undergraduates who majored in education and they were exposed to the Mastery Learning

Strategy for the statistics course, which involved the following phases:

- Presentation of group-based instruction.
- Administration of a diagnostic test.
- Identification of those students who did not reach a set standard on the test. These students were engaged in corrective activities.
- Students who were engaged in corrective activities were retested with another diagnostic test.
- Administration of summative test.

The major difference between the Mastery Learning Strategy statistics course and the lecture based course was the use of diagnostic tests and corrective activities in the former. The diagnostic tests were designed to pinpoint students' misconceptions. To complete these diagnostic tests, the students acted as diagnosticians because they had to identify whether the problems on the test had been solved incorrectly, describe the erroneous steps and propose appropriate solutions. A variety of statistical problems were used as the corrective activities in the study, which were designed to overcome the students' misconceptions. Each set of corrective activities contained four problems, each one of the following types: concrete-familiar, abstract-familiar, concrete-unfamiliar and abstract-unfamiliar. Here, a concrete statistical problem was defined as one which described a real situation dealing with real objects (e.g., salaries) and an abstract problem involved abstract data (e.g., algebraic symbols). A familiar statistical problem involved grades and an unfamiliar problem dealt with scientific ideas. Students did not simply attempt to solve these set problems, but were encouraged to predict the answers and to explain their solutions. In addition, students' misconceptions were made explicit to them and they were asked to explain the basis for their mistakes. After students had worked though the corrective activities they then completed another diagnostic test to ensure that they understood the concepts.

Students in both the experimental and control groups were tested after the course and the findings showed that students in the Mastery Learning Strategy group achieved significantly higher scores on these summative tests than the control group. It is, however, surprising to find that although the diagnostic tests emphasised students' understanding of statistical concepts, the summative tests simply consisted of multiple choice items (Mevarech, 1983). Student responses to multiple choice tests do not provide sufficient evidence that students understand the statistical concepts in question. However, Mevarech (1983) suggested that students must be engaged in corrective activities in order that statistical misconceptions are eradicated. Even if this suggestion is accepted it is still not clear whether the diagnostic tests or the corrective activities addressed students' statistical misconceptions.

The variety of empirical studies described above indicate that students are likely to have difficulties acquiring statistical concepts. It is worth noting that there is a meaningful sequence to these empirical studies: Mevarech (1983) looked into students' systematic errors in descriptive statistics in an attempt to extend the work of Pollatsek and his associates, but also because much research had been conducted on misconceptions regarding probability (e.g., Kahneman & Tversky, 1972). However, a number of issues that concern these studies can be raised. As has been previously noted, the majority of studies have been concerned with identifying students' misconceptions in statistics, but with the exception of Mevarech (1983), little research has looked at how such misconceptions can be remedied. It is noteworthy that the above range of studies have involved different participants in terms of their prior knowledge of statistics and their degree programme. For example, Pollatsek et al (1981) report that the seventeen participants in their study were undergraduate volunteers and that most of these (the number is not specified) were psychology majors and that three of the participants had completed "approximately half a semester of statistics" (p. 194). When research that concerns students' understanding of statistics is conducted, characteristics of the sample must be clearly detailed because such characteristics might have implications for the

research findings. It is striking that empirical work has used statistical tasks to assess students' understanding of particular concepts which are not representative of the kinds of problems that students have and will encounter in studying statistics: weighted mean problems are an example here. It is also important that research that looks at students' understanding of statistics should ask students to think aloud while they solve statistical problems so that valuable data concerning misconceptions can be collected. For example, Pollatsek et al (1981) asked students to think out aloud as they worked on a problem and also interviewed participants to determine the reasoning underlying their answers.

Students' misconceptions in statistics must be identified so that effective instructional materials can be designed and developed. Although much research has looked at the psychology of probability and some research has investigated students' understanding of the mean, there are other important areas of statistical enquiry that are of relevance to students taking psychology that need to be the subject of research. One such area, the focus of the research reported in this thesis, is correlation and how it might be understood by psychology students.

2.5 Correlation

The terms correlation and association are both used generally to refer to situations in which a relationship exists between variables. In the general sense, association is used to describe the degree of dependence or independence that exists between variables whether they are measured qualitatively or quantitatively. In the narrower sense, association is used to describe a relationship between dichotomous variables (Marriott, 1990). The term correlation is also used in a general manner to denote the interdependence that might exist between qualitative or quantitative data and in this sense includes the association of dichotomous variables and the contingency of classified attributes. However, the term correlation is more frequently used in the narrower sense to describe the relationship between measurable variates or ranks (Marriott, 1990). Different measures or coefficients of correlation and association can be used and the choice of coefficient will depend on the

type of data. This thesis is concerned with the narrower sense of the term correlation and the term association is only used when relevant research is discussed that has used this latter term (e.g., Batanero, Estepa & Godino, 1997).

Some texts aimed at the psychology student refer to correlational designs (Greene & d'Oliveira, 1982; Shavelson, 1981). Quite simply this is when a study is carried out to see whether there is a relationship between variables that do not readily lend themselves to direct experimental control or manipulation. In a correlational design, questionnaires are often used to see if certain variables are related. For example, in personality research, questionnaires have been typically used to study relationships between variables (Pervin, 1989). This situation is contrasted with the psychology experiment where, for example, an independent variable is manipulated to see if it has an effect on the dependent variable (Green & d'Oliveira, 1982). Correlational designs are often used in psychology because ethical and practical considerations prevent the direct manipulation of particular variables. As considered below, if a relationship is found to exist between variables then causal explanations for the relationship should not be inferred. However, students must learn about correlations because they are widely used in empirical work. If a correlation is obtained between two variables, this might suggest that further empirical work is required to look in detail at the possible nature of the relationship.

In introductory statistics courses correlation and regression are met in an investigation of bivariate data (Hawkins et al, 1992). Statistics textbooks that are aimed at psychology students include chapters that cover the topics correlation and regression (e.g., Hinton, 1995; Howell, 1992), whereas some texts contain separate chapters for these topics (e.g., Coolican, 1990; Pagano, 1990; Shavelson, 1981). Hawkins et al (1992) have noted a misconception that concerns the relationship between the techniques of correlation and regression: students view these two "topics as being two sides of the same coin" (p. 51). They are not. For example, a student might find that a high correlation indicates a strong relationship between two variables, and then goes on to

work out a regression of one variable on the other variable. There is, however, an important distinction to be made between a correlation and a regression analysis that would made this kind of thinking incorrect (Hawkins et al, 1992). Both correlation and regression are used in dealing with bivariate data, but to carry out a regression analysis one important assumption is made: one of the variables is treated as the response variable and the other as the explanatory variable (Daly, Hand, Jones, Lunn & McConway, 1995), and this explanatory or independent variable is fixed or specified by the researcher before the data are collected (Hawkins et al, 1992; Howell, 1992). In regression, the focus is on trying to explain how measurement on one variable changes in response to changes in the other variable. By contrast, correlation is used in situations when both of the variables are random (Howell, 1992). Although both regression and correlation are concerned with the relationship between two variables, regression is primarily used for prediction, whereas correlation is used to find out whether a relationship exists between two variables (Pagano, 1990):

> "Correlation is a topic that studies the relationship between two variables. Interest centres on the direction and the degree of the relationship" (ibid., p. 117).

Correlation is an important concept to be learnt by psychology students because techniques of correlation are widely used in psychological research. Coolican (1990) has outlined some of the common uses of correlation is psychology: reliability, factor analysis and twin studies. For example, correlation can establish inter-rater reliability of the judgment between people (raters). Factors analysis uses all the possible correlations between several tests taken by the same individuals, and IQ scores for twin pairs have been correlated (Coolican, 1990). Other uses of correlation in psychology abound: to see if the average age at which babies start to crawl is related to the average temperature for the sixth month following birth (Benson, 1993); to see if a relationship exists between TV violence and children's aggression (Eron, Huesman, Lefkowitz & Walder, 1972); a look at the relationship between stress and psychological symptomatology (Wagner, Compas & Howell, 1988); and to investigate if a relationship exists between an index of brain size

and people's IQ (Willermann, Schultz, Rutledge & Bigler, 1991).

The concept of correlation concerns both the direction and degree or magnitude of a relationship between two variables (Pagano, 1990). To understand correlation, a student must know that a relationship between two variables can be non existent, positive or negative. In addition, a student must learn that a relationship can vary from being nonexistent to being a perfect relationship (Pagano, 1990). There is often talk of the strength of a relationship as expressed by, for example, a correlation coefficient (e.g., Coolican, 1990). An important derivative of the concept of correlation is often described as follows:

"Correlation is *not* causation" (Daly et al, 1995, p. 438).

Texts emphasise that if a relationship is found to exist between two variables, then it cannot be concluded that there is a causal relationship between the two variables in question. Students (and researchers) must interpret a correlation with caution because although a correlation indicates that generally the values of two variables covary, there are four interpretations of an obtained correlation:

- One variable, A is the cause of the other variable, B.
- Variable B is the cause of variable A.
- The correlation between variables A and B is spurious. That is, the observed relationship is due to sampling variability or from sampling, for example, unusual behaviour.
- A third variable, C might be responsible for the obtained correlation between A and B.

(Coolican, 1990; Daly et al, 1995; Pagano, 1990).

When students learn about correlation they will use the main techniques of this topic which are the creation of scatter plots and the computation of correlation coefficients.

Bivariate data can be plotted on a scatter plot which will give an indication of the direction and strength of a relationship between two variables. A correlation coefficient provides a measure of both the kind of relationship that exists between two variables, and the degree of this relationship:

> "A correlation coefficient expresses quantitatively the magnitude and direction of the relationship" (Pagano, 1990, p. 118).

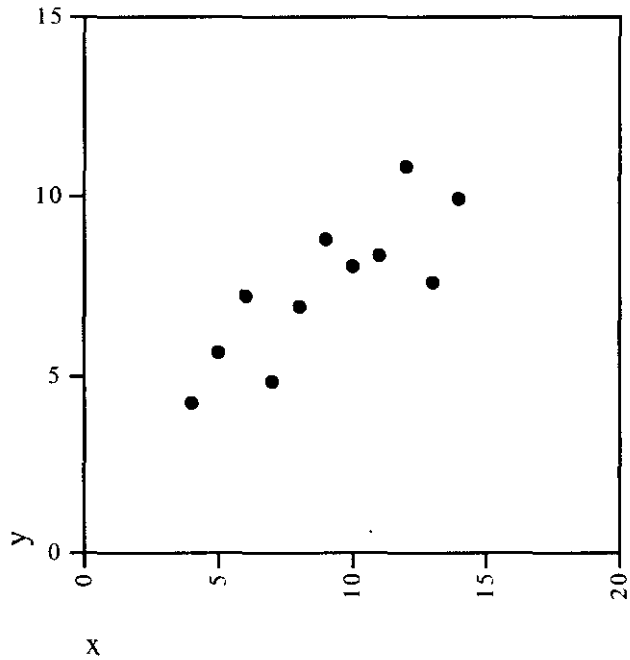
A correlation coefficient takes values between -1 (a perfect negative correlation) to 1 (a perfect positive correlation) and a correlation of zero indicates no relationship between variables (Daly et al, 1995; Pagano, 1990). Statistical texts for the behavioural sciences usually refer to the strength of a relationship as indicated by the magnitude of a correlation coefficient. The strength of relationship is expressed on a scale that ranges from -1 through zero to 1 as indicated by the value of the correlation coefficient. Coolican (1990) provides a scale, which describes in general, that irrespective of the sign, a correlation of between 0.1 and 0.3 indicates a weak relationship, a correlation of between 0.3 to 0.6 indicates a moderate relationship and a correlation of between 0.6 and 0.9 indicates a strong relationship. Although the statistical significance of a sample correlation is considered below, it is worth noting that if the sample size is fairly substantial, involving 30 participants, for example, a correlation of only 0.3 will be statistically significant ($p < 0.05$).

A number of correlation coefficients exist, but the Pearson correlation coefficient, which is denoted by r, is the most commonly used and is a measure of the strength of linear (straight line) correlation (Daly et al, 1995). The choice of the correlation coefficient will depend on the type of data where, for example, the Pearson correlation coefficient is used when data are measured on an interval scale, and the Spearman rank correlation coefficient is used when one or both of the variables are ordinal (Pagano, 1990). In addition to working out the value of a correlation coefficient, bivariate data should be described graphically because outliers can decrease the value of a correlation coefficient.

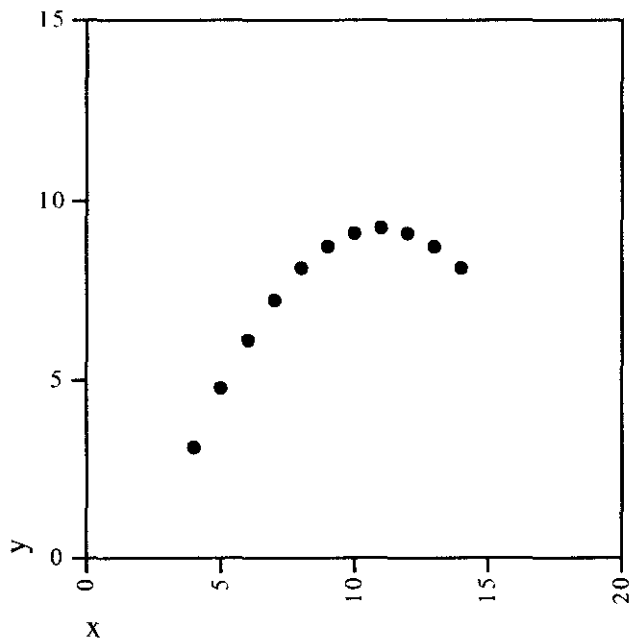
In addition, if data are curvilinear the Pearson correlation coefficient can underestimate the degree of relationship that exists between two variables (Pagano, 1990). Anscombe (1973) provided four different sets of artificial bivariate data all of which take a Pearson correlation coefficient of $r = 0.816$. The scatter plots for these data sets are shown in figure 2.1.

Figure 2.1 Scatter plots of data sets from Anscombe (1973)

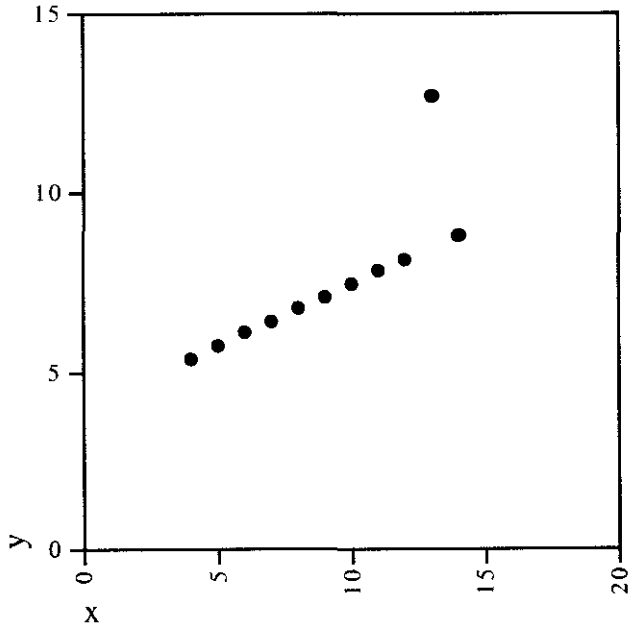
(a)



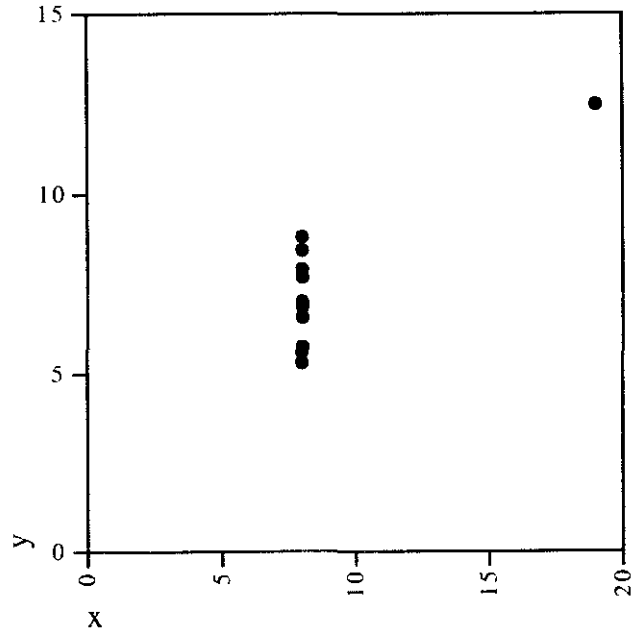
(b)



(c)



(d)



(Data sets from Anscombe, 1973).

In (a) the correlation 0.816 does provide a valuable summary of the bivariate data. However, in (b) the correlation does not provide an appropriate measure of the relationship because even though there is evidently a very strong (if not perfect) relationship between the variables, the relationship is not linear and for the Pearson correlation coefficient a perfect correlation (-1 or 1) is only found in a straight line relationship. Daly et al (1995) point out that in (c) the outlier decreases the value of the Pearson correlation, and if the outlier in (d) was excluded from the data set there would be no relationship between the variables. Such data emphasises the importance of graphical representation to determine if a relationship is linear before computing a correlation coefficient. Hinton (1995) points out that care must be taken to check that a data set has homoscedasticity when a correlation in being undertaken. Essentially, homoscedasticity means that the relationship between two variables remains the same at all points and that all the scores are evenly spread. 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 (Hinton, 1995). It is interesting that in Fox and Fowler's (1996) teaching experience students tend to blindly accept the calculated value of the correlation coefficient without considering graphical evidence.

One of the other most frequently used coefficients is the Spearman rank correlation coefficient (r_S) that is used for data that is ordinal (Coolican, 1990) and which provides a measure of monotonic correlation (Daly et al, 1995).

The question of interest is often not what the strength of relationship is between variables in a sample, but what the strength of relationship is in the population (Daly, 1995). The sample correlation coefficient is used to test the null hypothesis that, in the population, there is no correlation or that the coefficient is zero (Daly, 1995). In psychology a correlation is usually tested (with a given sample size) to see if it is significant at a level of probability. Typically, statistics texts have provided statistical

tables so that a student can check to see if a correlation is significant at the level of 0.01 or 0.05. For example, a student would find that a correlation of 0.79, for a sample size of 12, is significant ($p < 0.01$). However, with the increasing use of data-analysis programs, such as *SPSS*, students might not continue to use statistical tables because such applications provide output that specifies the probability of an obtained statistic.

In summary then, when psychology students study the topic of correlation, they will learn about correlational designs, positive and negative correlations and will address the issue of causation. In addition, they will employ the primary techniques of correlation: the production of scatter plots and the calculation of correlation coefficients. Students will learn how to test a correlation to see if it is statistically significant and must also be able to interpret the meaning of correlation coefficients when they are, for example, reported in the psychology research literature. There has been very little research that has looked at the kinds of difficulties that students encounter in these areas. In addition, no empirical work has been found that has looked at the relationship between students' general mathematical understanding and students' understanding of the topic of correlation. There has, however, been some research that has focused on people's detection and assessment of covariation.

2.6 The assessment of covariation

Chapman and Chapman (1982) have described a phenomenon termed illusory correlation where there is a "tendency to see two things as occurring together more often than they actually do" (p. 241). L. Chapman (1967) found that semantically related words tended to be seen as occurring together in pairs more often than they really did. In a study various word pairs, such as bacon-tiger were projected onto a screen in front of a participant. These word pairs were changed every two seconds and one of the words of the pair was always one of four possible words (bacon, lion, blossoms, boat). Each of these words appeared for an equal amount of time and were paired with the words: eggs, tiger or notebook. The word pairs were systematically arranged so that each left side

word appeared an equal number of times with each right side word. However, when participants were asked about the word pairs, they said, for example, that when the word bacon appeared the word eggs was paired with it 47 per cent of the time. In general, participants claimed that the word pairs with a verbal association occurred more often than the other word pairs even though every word pair were presented to participants as often as every other (Chapman 1967). With regard to students learning correlation as a statistical topic, the implications of such findings are unclear. Illusory correlation is vaguely described and Chapman and Chapman (1982) have used the term correlation generally to refer to "two things" that go together. For students correlation is likely to mean much more than this: they typically have to deal with bivariate data measured on an interval or ordinal scale and must learn to employ techniques of correlation to assess whether a relationship exists between two variables and to interpret a possible relationship (or lack of it).

Research has also been concerned with lay persons' proficiency at the detection and assessment of covariation (Jennings, Amabile & Ross, 1982; Well et al, 1988). Jennings et al's (1982) study is considered here in detail because although it has been cited as relevant in the context of research into statistical education (Hawkins et al, 1992), there are a number of criticisms that can be raised about this study. Jennings et al (1982) were concerned with people's difficulties at detecting and assessing covariation and in the light of this raise two important issues of interest that they investigated in an empirical study. Firstly, they looked at people's proficiency at data-based assessment of covariation. That is, as shall be considered below, they investigated people's difficulties in detecting and assessing covariate data that were "unencumbered" by theories about the empirical relationship (Jennings et al, 1982, p. 216). Secondly, they investigated whether there was a tendency for people's theories (or expectations) to lead them to assume a strong correlation when no empirical evidence was available. Two different kinds of tasks were therefore used to investigate participants' covariation assessment. The first kind of task was used for participants' data-based assessments of covariation and also provided data

about the relationship between subjective judgments of covariation and objective measures of covariation. Here, participants were presented with sets of bivariate observations that they could apparently hold no theories or expectations about and were asked to assess the strength of the relationship in each set. More specifically, participants were provided with, for example, a set of 10 simple number pairs, which they were asked to study and estimate the relationship between by placing an 'X' on a 100-point subjective rating scale with two end points that read 'perfect relationship' and 'no relationship'. The second kind of task was used for a theory-based assessment of covariation. Here, pairs of variables were used that apparently would be linked by people's theories. Participants were asked to estimate first the direction and then the strength of the relationship between these variables by using the 100-point subjective rating scale. For example, for the second kind of task, participants were presented with the following two pairs of variables:

Children's dishonesty as measured by false report of athletic performance.

Children's dishonesty as measured by amount of cheating in solving a puzzle

[$r = 0.18$]

Students' height.

Students' weight

[$r = 0.79$].

(Jennings et al, 1982, p. 218).

Such pairs of variables were presented to participants, but the correlation coefficient obtained for each pair was not presented. However, the correlations were selected from previous empirical studies and therefore provided an objective measure of the relationship between the presented pairs of variables. These objective measures of correlation were required so that a participant's subjective assessment of the relationship between pairs of variables could be compared to an objective measure and assessed for accuracy (Jennings

et al, 1982).

Participants' performance at data-based covariation estimates provided an indication of how easily the participants detected various degrees of covariation in the absence of any theory about the relevant bivariate data. Jennings et al (1982) found that participants found the data-based estimation task difficult. For example, many participants had difficulty recognising positive relationships that are commonly reported in psychological research (e.g., 0.4). Such correlations were barely detected by participants who gave mean estimates for such correlations in the range of only four to eight on the 100-point scale. Even strong positive correlations such as 0.6 resulted in low subjective estimates of covariation. For instance, an objective correlation of 0.7 produced a mean subjective rating estimate of 35, a rating that was midway between the points on the rating scale labelled 'rather weak' and 'moderate' (Jennings et al, 1982).

In the case of theory-based estimates, there was a tendency for participants' estimates to show an approximate correspondence with the objective measures. So, for example, positive relationships were estimated to be positive and relatively strong relationships were estimated to be stronger than relatively weak ones (Jennings et al, 1982). However, participants were not as conservative in their theory-based estimates of covariation as they were in their data-based ones. For example, the following pair of variables was presented to participants:

Sixth graders' ability to 'delay gratification'

Sixth graders' ability to resist temptation to cheat.

[$r = 0.31$]

(Jennings et al, 1982, p. 218).

When presented with bivariate data that would give a correlation within the range $r = 0.2$ to $r = 0.3$ participants provided estimates of between zero and 10 on the rating scale, but given this pair of variables (with no data or coefficient) participants' theory-based

estimates averaged in the 50 to 60 range of the 100-point rating scale (Jennings et al, 1982).

A number of criticisms concerning the methods used in the above study can be made. Firstly, it is unrealistic to expect that people can provide a reasonable estimate of covariation based on the kinds of stimulus materials used in the data-based estimation task. Three sets of stimulus materials were used to provide bivariate observations: sets of 10 number pairs; drawings, each giving a set of 10 men of various heights holding walking sticks of various lengths; and presented in the form of audio a set of 10 individuals were heard to emit some letter (that occupied an ordinal position in the alphabet) and then sing a musical note of varying duration. These materials do not provide an adequate set of bivariate data. Although Jennings et al (1982) were concerned with lay persons' detection and assessment of covariation (the study involved 64 undergraduates who had not taken a college level course in statistics) it would have been reasonable to present scatter plots of bivariate data and to ask participants to estimate both the kind and magnitude of the relationship. The empirical study was concerned with lay persons' assessment of covariation, but the pairs of variables used for the theory-based assessments were not appropriate for the lay person who might not be aware of the complexities of psychological research. Take the following pair of variables that was presented to participants:

Self-ratings of political conservatism

A composite of self-rating items from the Ethnocentricity scale

(Jennings et al, 1982, p. 218).

What possible theory or expectation could a lay person have about the relationship between these two variables? On what basis would they be expected to provide a reasonable estimate of the relationship that takes an objective correlation of 0.57? It is doubtful whether a lay person holds a theory that concerns the Ethnocentricity scale and

what variable(s) it might be linked to.

As well as the above criticisms, Jennings et al (1982) use the findings of the study to make unjustified generalisations:

> "When no objective, ... bivariate data can be examined, but prior theories or preconceptions can be brought to bear, the intuitive psychologist is apt to expect and *predict* covariations of considerable magnitude - often of far greater magnitude than are likely to be presented by past experience or to be bourne out by future experience" (my emphasis, ibid., p. 224).

The participants of the above study were not asked about their expectations or theories concerning the relationships between the pairs of presented variables and they were not asked to predict covariation: participants were asked to provide a subjective rating to estimate the relationship between the variables. It is noteworthy that research (Well et al, 1988) has suggested that a 100-point subjective rating scale is not necessarily a reliable tool to use for participants' assessment of covariation.

It has been proposed that Jennings et al's (1982) findings might be dependent on the format of data presentation used in the research (i.e., sets of digit pairs in data-based estimates) (Lane, Anderson & Kellam, 1985). Research has shown that people perceive variables to be more highly related if the data are presented graphically than if they are presented in a tabular format (Lane et al, 1985). In one of a series of experiments by Lane et al, participants, who were undergraduate psychology students, were presented with tables of bivariate data or graphs showing scatter plots of the same data and were asked to detect and estimate the relationship between variables in the tables or the scatter plots (Lane et al, 1985). Participants were asked to measure the relationship on a scale of zero to 100, where they were told that zero meant no relationship and 100 meant a perfect linear relationship. Overall, participants judged the degree of relatedness between the continuous variables to be higher in the graphical format than in the tabular format. Another experiment, which made use of the same method and stimulus materials,

involved participants who made up a statistically expert sample. The participants were professionals in psychology, economics and education who had volunteered themselves for the study as part of a statistically expert sample and who had taken courses in statistics at college. It was found that these experts tended to see stronger relationships in data presented in a tabular format than did novices (Lane et al, 1985).

Well and his associates (1988) also conducted a study designed to investigate people's assessment of covariation of continuous variables. In this study, however, participants were asked to estimate three sets of paired two digits numbers (each set consisting of 60 X-Y pairings!) that had the correlations of 0.9, 0.6 and 0.1. These researchers also made use of the familiar rating scale that was labelled zero at one end to indicate no relationship and 100 at the other end to indicate a perfect relationship. This was used by the participants to provide an estimate of the strength of the relationship between the sets of X-Y pairings. It was found that the participants, who were undergraduate students that took psychology courses, were quite poor at providing estimates of the strength of relationships (Well et al, 1988). However, these researchers found that there was variability in how participants used the rating scale where the judgment range (the difference between participants' largest and smallest estimate) varied considerably between participants. This calls into question the use of a subjective rating scale as a reliable and valuable measure for people's estimate of covariation.

Research studies that have looked at people's detection and assessment of covariation have not used bivariate data that had negative correlations (Lane et al, 1985; Well et al, 1988). Jennings et al's (1982) research did include pairs of variables that had objective measures of negative correlations, but it is not at all clear whether any of the bivariate observations used in the study had negative correlations. In a study that investigated people's estimation of correlation from scatter plots, negative correlations were also not used (Strahan & Hansen, 1978). Strahan and Hansen (1978) recruited participants who were either enrolled in an advanced basic statistics course and were

predominantly graduate students or were faculty from a psychology department. In this study, participants were asked to estimate the degree of relationship from scatter plots by providing a correlation coefficient with two decimal places. Strahan and Hansen (1978) found that participants underestimated the degree of correlation represented by scatter plots. Strahan and Hansen (1978) noted that a limitation of their research is that the entire range of correlations from -1 to 1 was not presented to participants and that this was to prevent possible confusion with negative numbers. Research is required to investigate students' assessment of bivariate data that shows a negative relationship.

The above studies have been primarily concerned with lay persons' assessment of covariation, with the exceptions of Strahan and Hansen's (1978) work and experiments that have looked at experts' and naive participants' assessment of bivariate data in both tabular and graphical formats (Lane et al, 1985). However, the statistically naive participants involved in Lane and his colleagues' research were undergraduate psychology students and although Well et al (1988) were concerned with assessment of covariation by people in general, the participants involved in this study were also undergraduate psychology students. In both these cases, the participants' prior knowledge of statistics is not reported.

It is difficult to say whether the above studies indicate that people in general find covariation assessment difficult. Jennings et al's (1982) study indicated that in the case of data-based assessment of covariation participants did not tend to detect positive correlations. However, this finding might be dependent on the format of data presentation that was used in the study and the use of a subjective rating scale that might not be reliable. Related research has indicated that people are quite poor at estimating the strength of relationship (Well et al, 1988), but to assess sensitivity to covariation detection this research only used pairs of numbers as bivariate data. Research that has looked at participants' estimates of the degree of correlations from scatter plots has shown that people tend to underestimate the relationship between variables (Strahan & Hansen,

1978). It is important to emphasise, however, that the purpose of the above studies has not been to identify the confusion and difficulties that psychology students have concerning correlation. It is somewhat incidental that at least two of the studies that concern people's assessment of covariation have recruited psychology students (Lane et al, 1985; Well et al, 1988).

When students study the topic of correlation they use particular techniques to establish whether a relationship exists between two variables. Accordingly, they must learn how to estimate the direction and magnitude of the relationship from a scatter plot, and to interpret the correlation coefficient obtained. Research has shown that students do have difficulties and confusions pertaining to these aspects of learning. With the exception of one recent research programme that is considered in the next chapter (Batanero, Estepa, Godino & Green, 1996), there has been a lack of research that has looked at students' understanding of correlation.

2.7 Summary

This chapter has examined research that relates to statistical misconceptions. The extent or prevalence of these misconceptions is not clear from this research literature. Indeed, this would be very difficult to determine because the characteristics of the participants in the studies reviewed differ in terms of, for example, whether they have previously studied statistics and this kind of prior knowledge might affect research findings relating to whether the participants held particular misconceptions.

One aspect of the research described in this thesis is how the topic of correlation might be understood by psychology students. Accordingly, the topic of correlation was outlined. Students taking psychology will learn about: correlational designs, the concept of correlation, causation and correlation, the techniques of correlation and how to test a correlation to see if it is significant.

Research that has concerned people's detection and assessment of covariation was examined. From such research it can be concluded that people do not tend to detect empirical correlations (Jennings et al, 1982) and are poor at estimating the strength of relationship (Well et al, 1988). It has also been demonstrated that people perceived variables to be more highly related if data is presented in a graphical format than if it is presented in a tabular format (Lane et al, 1985), but that people tend to underestimate the magnitude of correlation from scatter plots (Strahan & Hansen, 1978).

This chapter has shown that the purpose of much of the research that relates to students' difficulties with correlation has been concerned with lay persons' assessment of covariation and not, for example, identifying particular confusions that psychology students might hold about correlation. There is very little research that has investigated university students' confusions and difficulties that concern correlation. Moreover, there is a lack of research that has looked specifically at psychology students' misconceptions concerning correlation. A recent research programme has looked at students' understanding of association and correlation (Batanero et al, 1996; Batanero et al, 1997; Estepa & Batanero, 1996). However, Batanero and her colleagues (1997) have been concerned with students' understanding of association in general and have conducted empirical work that has looked at students' conceptions about association in contingency tables (Batanero et al, 1996) and at students' conceptions about correlation in scatter plots (Estepa & Batanero, 1996). This research programme has not looked at psychology students' misconceptions concerning correlation, but has investigated pre-university students (17-18 year old students) understanding of association (Batanero et al, 1996; Estepa & Batanero, 1996). These students had not received any teaching of statistical association and the research therefore concerned students' preconceptions concerning statistical association (and correlation). Similarly, further experimental work involved 19-20 year old university students who had not previously studied statistical association (Batanero et al, 1997). This research programme is very recent and could not therefore

inform the design of the investigation that is outlined in the next chapter. Batanero et al's (ibid.) empirical work is, however, discussed in relation to the findings of this investigation in chapter 3.

In contrast to Batanero and her associates' work, the research reported in this thesis specifically investigated university psychology students' confusions and difficulties concerning correlation. In addition, the research was designed to involve students who had already studied correlation or who had received instruction in this area because it is possible that, for example, by learning about correlation students acquire particular statistical misconceptions. As outlined in chapter 1, with the advances in computer technology, computer-assisted learning programs are likely to be used increasingly in higher education and could provide an additional form of instruction to help students acquire statistical concepts. However, the effective design of such programs should be informed by both research-based principles of learning and empirical work that looks at students' understanding of statistical topics (chapter 4). With this in mind, the empirical study that is described in the following chapter was conducted.

Chapter 3

An investigation of students' conceptions and skills pertaining to the statistical topic of correlation

3.1 Introduction

The previous chapter indicated that students are likely to find statistics difficult, but there has been a lack of research that has looked at the kinds of confusions and difficulties that psychology students have in the particular area of correlation. This chapter describes a study that was conducted to investigate whether students find statistics difficult and to identify students' confusions and difficulties pertaining to the statistical topic of correlation.

3.2 Methodology for the investigation

One of the purposes of the study was to investigate students' experience in studying statistics in terms of interest, difficulty and enjoyment, and also how they would describe statistics as a subject area to study. It has been suggested that typically students find statistics difficult (Shute & Gawlick-Grendell, 1993), but these researchers have not provided empirical evidence to support this view.

Drawing on research concerning students' misconceptions, it was suggested in the previous chapter that research concerning students' understanding of statistical topics should employ techniques that allow light to be shed on students' thinking. If a student simply provides the correct answer to a statistical task or problem this does not necessarily mean that they fully understand the statistical concept in question.

Empirical studies on statistical understanding have collected data concerning students' thinking or reasoning by asking students to think aloud while they have attempted to answer certain questions (e.g., Konold, 1989; Pollatsek et al, 1981).

Ericsson and Simon (1980, 1993) have maintained that concurrent verbalisation, in which the participant is asked to think aloud while they work on a task, provides valuable data on participants' thoughts. Indeed, concurrent verbal reports are now generally accepted as major sources of data on people's cognitive processes on particular tasks (Ericsson & Simon, 1993). Ericsson and Simon (1993) have recommended guidelines for the use of think-aloud: for example, warm-up exercises should be employed in a procedure that asks students to think aloud so that participants can practice verbalisation of their thoughts. In the investigation that is described in this chapter, participants were asked to think aloud while they completed the statistical tasks and warm up exercises were provided. It has also been recommended that social interaction between the participant working on the task and the researcher is minimised and that, if necessary, the researcher should only remind participants to think aloud (Ericsson & Simon, 1993). However, the investigation, which is described in this chapter, required a method that prescribed the use of prompts. This was necessary because while students worked on the statistical tasks a method of prompting was needed for when students became stuck on a particular task and the use of predefined prompts meant that data concerning a student's particular difficulty, for example, could be collected. The method in the study used a predefined plan, which is described below, to prompt students if they got stuck on a particular statistical problem.

To find out more about students' understanding of particular statistical concepts statistical tasks or problems are required for students to complete. There is a lack of available research instruments or tasks that can be used in empirical work to assess students' understanding of correlation. This is probably because there has been little research that has looked into students' understanding in this area. A variety of tasks or questions were developed for the investigation that is described in this chapter. It was decided that the questions should be piloted to see which particular questions were useful in terms of testing students' understanding of particular concepts that concern correlation. So, an additional purpose of the study was to devise and pilot tasks for the topic area of correlation.

Konold (1989) investigated students' conceptions of probability by setting a series of problems for them to answer and asking students to think aloud while they attempted to solve each problem. In other words, Konold (1989) interviewed students individually while they attempted to verbalise their thoughts as they occurred. Probes that were used during the interviews consisted of requests to repeat a comment and reminders to think aloud, and unplanned probes were also used in an attempt to clarify students' thinking.

In a study that made use of clinical interviews, Perkins and Martin (1986) watched students individually as they solved problems in programming and occasionally asked questions to track a student's thinking until they encountered a particular difficulty or simply got stuck on the task at hand. The researcher observed and interacted in defined ways with the students as they worked through problems (Perkins & Martin, 1986). To track a student's thinking and to provide help while a student worked through the tasks, the researcher, if it was necessary, progressively moved from the use of general prompts to hints and finally, to the use of what are termed 'provides' to resolve a student's difficulties. For example, if a student got stuck they would be initially prompted by "high-level strategic questions one might ask oneself" (ibid., p. 216), which would include, for example, "what is your plan?" or "how would you describe the problem to yourself?" If prompts did not promote progress on the problem then the researcher would move to the use of hints that reflected the researcher's understanding of the solution to the problem and, if this did not help, to the use of provides. For a provide, a student would be given the exact solution to the specific difficulty at hand. This approach provides a plan of what to do if the student gets stuck for an empirical study that involves observing students while they work on tasks. This approach was modified and employed in the investigation that is described in this chapter. Primarily, general prompts were used while students worked through the statistical tasks, and if necessary hints and provides were used for a question (question 14), which asked students to calculate a statistic by using a formula.

A further purpose of the study was to investigate students' conceptual difficulties and procedural skills in the topic of correlation. As previously noted in chapter 2, it has been suggested that students find statistics difficult because they have inadequacies in prerequisite mathematical skills (Garfield & Ahlgren, 1988). If this is the case, then it is likely that students will also experience difficulties in carrying out particular statistical procedures or they will lack the necessary skills required to derive a procedure from a formula so that they can calculate a statistic.

It is likely that students will have prior conceptions when they begin to take a course in statistics that might interfere with the acquisition of statistical concepts concerning correlation. However, the target population for this study were students taking psychology at an institute of higher education who had already taken a majority of the required statistics courses for their degree programme. This meant that the kinds of tasks, which were developed to investigate students' understanding and difficulties concerning correlation, were not so restricted in their scope and detail. It would be tricky to ask a student to produce a scatter plot from data if they had very little or no prior knowledge of how to do this.

To reiterate, the four main objectives of the study were to investigate:

(i) Students' subjective opinions about whether they find statistics difficult

(ii) The kinds of tasks that can be used to assess students' understanding of correlation

(iii) Students' conceptions concerning correlation

(iv) Whether students have difficulties in carrying out particular statistical procedures

3.3 Method

3.3.1 Participants

Twenty students, (thirteen females and seven males) four of whom were from the University of Buckingham and sixteen who were from the University of Luton, took part in the study. They were paid £3 per hour for their participation. The mean age of these students was 23 years (minimum age 19, maximum age 37). Eleven of the students were in the second year of their undergraduate degree programmes and nine of them were in their final year. Eleven students were taking a degree in psychology and nine were taking a joint degree that included psychology. All of the students had completed an undergraduate course in statistics called Research and Experimentation, which covered correlations. Two students had obtained an A' level in mathematics and statistics and one student had an equivalent qualification to A' level mathematics called Matric Mathematics. Fifteen of the students had obtained a grade C or above in GCSE mathematics or had an O' level in mathematics. One student did not have a GCSE or an O' level in mathematics or statistics (or an equivalent), but she had an HNC that involved Business and Finance. To ensure anonymity (id1) for example, refers to a participant.

3.3.2 Materials

Information sheet

This was used to inform students of the purpose of the study and what it involved (Appendix A). For example, this sheet made it clear that if students got stuck on a particular task, there were a series of steps that would be followed to help them to continue. In addition, this sheet told students that they were to complete a task booklet at their own pace and to think aloud while they worked. Students were also asked to save any questions that might come to mind while they were working to the end of the session.

Questionnaire

A questionnaire was devised and used to collect data concerning student characteristics and details of their degree course and institution. Here, students were also asked to specify any formal qualifications that they had in mathematics, and the materials and texts that they used for statistics. In addition, the questionnaire was designed to ascertain whether the students found statistics difficult.

Pre-task sheet

This sheet included warm-up exercises that were used initially to make sure that the students became familiar with the idea of thinking aloud before they worked through the task booklet.

Task booklet

The task booklet was devised to include questions that related to particular subtopics of correlation. To identify those subtopics that are typically covered the following texts were consulted: Greene and d'Oliveira (1982), Hinton (1995) and The Open University (1990). For each subtopic, these texts were also used to devise a variety of tasks in the form of questions and appropriate answers to these questions. One of the questions was based on a study that was described in Smith and Cowie (1988) and originally described in Eron et al (1972). The choice and organisation of the subtopics, and the questions and their answers were examined and checked by three statistics specialists to ensure that they were accurate, nonambigious and appropriate. The following subtopics were identified as comprising the topic area of correlation:

- Correlational designs
- Scatter plots
- Positive correlation
- Negative correlation

- Zero correlation
- The strength of a correlation
- The null hypothesis
- Parametric test: Pearson product moment correlation coefficient
- Parametric test: the significance of a correlation coefficient
- Interpretation of data

A task booklet for the students to complete was therefore developed, which consisted of questions that involved graphs and data, and space for the students' answers. This booklet was designed to take a student approximately one hour to complete. A standard task booklet was also developed to include both the questions in statistics and the model answers to the questions (appendix B). The students' answers to the questions were then categorised with reference to the model answers that were devised.

Audio cassette recorder

This was used to record the students' comments as they completed the questionnaire and the students' think-aloud as they worked through the task booklet.

Prompting sheet

A sheet was used that consisted of prompts that were generally applied when the students worked through the booklet. These kinds of prompts included the following:

What are you thinking?
Remember to say out-loud what you are thinking.
Write down what you think.
I can tell you that at the end.
I can show you that at the end.
If you like, move on to the next question.

This sheet also contained prompts that were used if the students got stuck on a particular question. For example, if a student could not answer a question they were asked: "what's the first thing that you need to do?" or "what is your plan?" or "what do you need to do next?" If prompts did not help the student while they attempted to answer question 14, which asked students to calculate a statistic by using a formula, hints were then used that were geared to the question, and if appropriate, provides were used to give a solution to an immediate specific problem. To answer question 14 students were required to derive a procedure from a formula and here, for example, a hint that was used suggested that the student make a column of data to calculate ($\sum a^2$) and a provide that was used involved telling the student that $\sum$ meant 'the sum of'.

Data recording sheet

This sheet was designed to reflect the structure of the task booklet and was used to supplement the think-aloud protocols and to record pertinent student comments relating to particular questions.

Study materials

A scientific calculator was available for the students as they worked through the questions. The students were not told how to use the statistics mode on this calculator, which calculated basic descriptive statistics, and none of the students used this mode.

3.3.3 Procedure

Firstly, students were provided with information about what the session would involve (the information sheet) and were asked to think aloud while they worked through the questions. They were then asked to complete the questionnaire. Secondly, to practice thinking aloud the students were asked to complete the pre-task sheet. Thirdly, students worked through the task booklet and used the spaces provided for their answers. Additional notes were taken of students' responses and comments while they worked

through this booklet. Students were prompted, if it was necessary, to clarify any of their answers, to remind them to think aloud, or to help them if they were stuck on a particular question. Finally, student queries concerning the session or problems relating to specific questions were dealt with.

3.4 Findings

The findings of the study are described in relation to the four main objectives of the study.

3.4.1 Statistics as difficult

The questionnaire asked the students two related questions: "What has been your experience in studying statistics in terms of interest, difficulty, enjoyment, etc.?" and "In your experience, how would you describe statistics as a subject area to study?" Each of the student's answers to both these questions were combined for analysis. To see whether students find statistics difficult, students' answers to both questions were categorised as either describing the subject as difficult, or not difficult. So, for example, one student's response to the first question above was to comment "extremely difficult" and to the second question above she wrote "certainly not easy" (id1). This student's answers were therefore categorised as describing the subject as difficult. Twelve of the students (60%) described statistics as difficult in some form or another. In response to the first question above, one student simply wrote "I have found statistics a difficult area of study" (id2) and another student commented "I don't enjoy doing the statistics part of the course. I find it difficult, with all the different terms and techniques" (id9), and another responded "I found it difficult to study" (id20).

A number of students, however, talked about their interest in the subject. The students' answers were also categorised in terms of whether they described statistics as interesting. Seven of the students (35%) described statistics as interesting. When asked to describe statistics as a subject area to study one student wrote "useful and quite

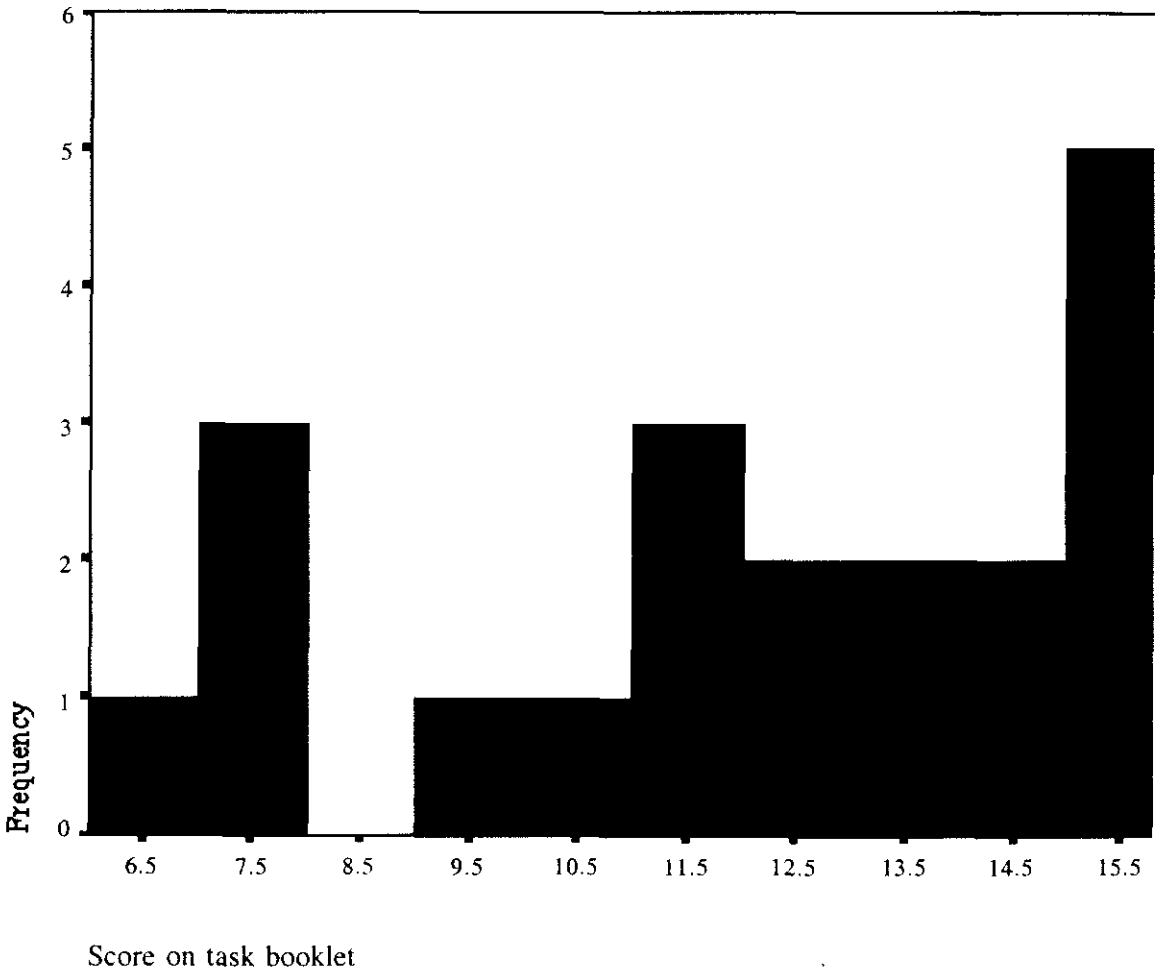
interesting" (id10) and another commented "I have found it an interesting area" (id16). In contrast, 20 per cent of the students described how they did not find statistics interesting. For example, one student wrote that they had "no enjoyment or interest in them - just do stats because we have to as part of the course" (id13). One student provided a reason as to why statistics might not be interesting: "most likely it would be interesting if you could understand it. Unfortunately I can't" (id12).

3.4.2 Task booklet

Students were asked on the questionnaire which texts they had used for the statistics courses they had taken. It was found that seventeen of the students (85%) used texts that had been consulted to devise questions for the task booklet: these participants used Hinton (1995) and Green and d'Oliveira (1982).

The students' answers to each of the questions that comprised the task booklet were initially categorised as appropriate or not. This meant that an overall score for the task booklet could be calculated for each student. The maximum possible score that could be obtained on the booklet was 19. This score provided an estimation of a student's understanding of correlation and an assessment of their skills in this topic. For the twenty students, the mean score for the task booklet was 11.7 (maximum = 16, minimum = 6, S.D. = 3.1). Figure 3.1 shows a histogram of the students' overall scores on the task booklet.

Figure 3.1 Histogram of students' scores on the task booklet



With regard to the overall scores on the task booklet, there was no significant difference in the means for those students who described statistics as difficult and for those students who did not (two-tailed test, $t = 1.15$, d.f. = 18, $p > .05$). Likewise, there was no significant difference in the means for the task booklet scores for those students who described statistics as interesting and for those that did not (two-tailed test, $t = 1.45$, d.f. = 18, $p > 0.05$).

For those answers that were inappropriate, 15 categories were created to account for the variety of student answers that were found. A category described an error or

misconception common to three or more students. These categories included ones that were question independent such as, 'lacks knowledge' in which the student, for example, responded to a question by writing that they did not understand the question, and categories that were question dependent. The generation and definition of categories for the students' answers was an iterative process. Table 3.1 summarises the findings concerning the task booklet. This table provides the answer categories of student responses for those questions that were valuable in that they highlighted particular misconceptions that students held.

Table 3.1 Answer categories for questions: per cent of student responses for particular categories

Question	Appropriate answer	Misconceptions	Other categories
Question 4	55% (11) [a]	• Causalistic 25% (5)	20% (4)
Question 8	50% (10)	• Negative viewed as no relationship 15% (3) • Negative viewed as positive 15% (3)	20% (4)
Question 10	60% (12)	• No correlation viewed as negative 15% (3)	25% (5)
Question 12	65% (13)	• Positive correlation stronger than negative 10% (2) • Positive correlation stronger than negative and no correlation viewed as negative 15% (3)	10% (2)
Question 17i	20% (4)	• Causalistic 20% (4)	60% (12)

[a] Frequencies are given in brackets.

Students' answers to each of the questions in the task booklet are now considered.

3.4.3 Students' answers to the task booklet

Question 1. In psychology, when would you use a correlational design?

In response to this question, it was anticipated that students should simply answer that you would use a correlational design if you wanted to see if there was a relationship or

correlation between two variables. Just over half of the students (55 per cent) answered the question in this way as one student wrote "when you want to test for a relationship between variables" (id1) and another student commented "when you want to see if there is a relationship between two variables" (id7). However, some of the students' answers to this question indicated that they were not clear about the nature of a correlational design. Twenty per cent of the answers were categorised as 'not correlational design' because students indicated that they did not have a clear conception of such a design. For example, one student answered that you would use a correlational design "when looking for a difference between 2 variables" (id19). It is worth noting that for one student this question uncovered the misconception that one is looking for causality in a correlational design because they commented "trying to find a cause and effect relationship between 2 variables" (id17).

Question 2. Give an example of a study that would make use of a correlational design.

For a study that makes use of a correlational design a researcher wants to see if a relationship exists between two variables that do not easily lend themselves to experimental control and manipulation. Typically, tests and questionnaires are used in this kind of design. Fifty five per cent of the students provided an appropriate example of a study that would use a correlational design where, for example, one student simply wrote in response to this question "heart disease and butter" (id4) and another wrote "smoking and cancer"(id14). Students were prompted if their answer was not sufficient in terms of the detail provided. One student wrote "bullying and self-esteem" and they were then prompted:

> E: "Can you say a little more?" ...
>
> S: "Well, the reason I said this is because this is what we're doing at the moment"
>
> E: "Yeah."
>
> S: "... You know, if whether you were bullied .... is related to self-esteem"
>
> (id12). (Excerpt from audio recording of session).

Fifteen per cent of the students' answers indicated that they did not have a clear conception of a correlational design because the examples they provided were inappropriate. Indeed, some of these students appeared to describe the design of an experiment, as one student wrote "if you wanted to see if drinking alcohol before going to sleep caused nightmares, you would have 2 groups of subjects, one who had alcohol and one who did not. You would then measure the number of nightmares in some way, at the same time control all extraneous/confounded variables" (id1).

Question 3. Let's suppose that a large-scale research study has reported that a significant correlation had been found between clinical depression and cancer. What do the findings tell us about the statistical relationship between clinical depression and cancer?

This question was designed to see if students inferred causality from correlation. Fifteen per cent of the students spoke of some kind of causal relationship that could exist between depression and cancer: one student wrote "it tells us that clinical depression can cause cancer but it's not necessarily the only variable" (id17). Seventy per cent of the students answered this question appropriately and thirty per cent of these students assumed that the statistical relationship between depression and cancer was positive, which is a reasonable interpretation of the question. For example, one student stated "could be that if you are clinically depressed you're more likely to suffer from cancer or conversely, if you have cancer, you're more likely to be clinically depressed" (id1) or, as another student wrote "that if a person is diagnosed as having clinical depression they are more likely to have cancer" (id3). However, it is not explicit in the question whether the statistical relationship between the variables is positive or negative. In this respect, some students were cautious in their interpretation: one student responded to this question by writing "there's a positive relationship between depression and cancer. I assumed significant to be a positive relationship as opposed to a negative one" (id9). Another student could not even make this assumption: "there is not a causal link between depression and cancer,

simply some sort of relationship between the two, although we don't know what that relationship is because the question doesn't tell us" (id13).

Question 4. If a correlational study finds a relationship between two variables, could you ever conclude that there is a causal relationship between two variables?

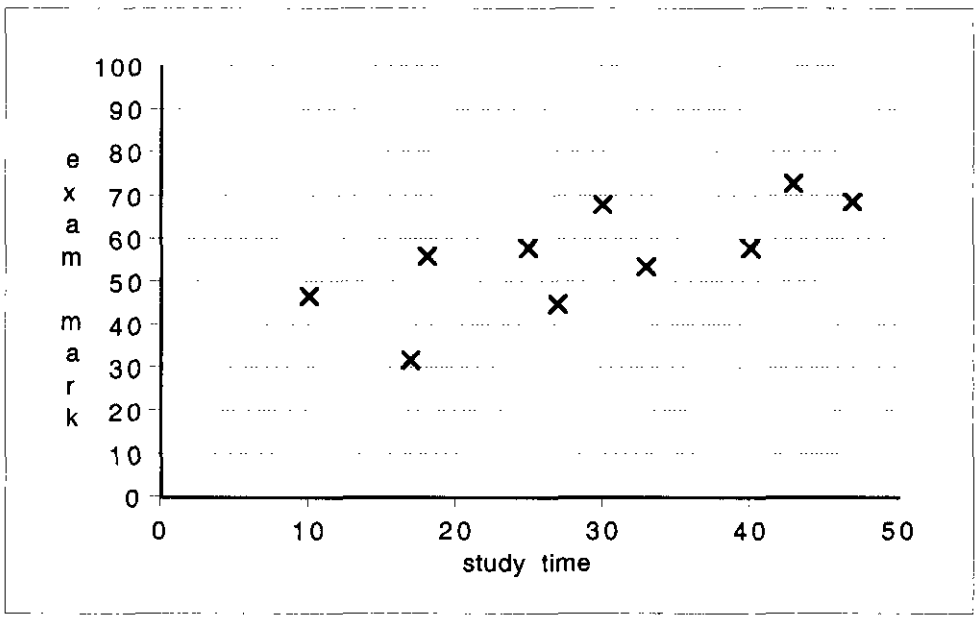
Although a student should answer no to this question, 25 per cent of the students commented in some form that one could infer causality from correlation. For example, one student who held this conception, wrote "if the population your sample was drawn from was large enough, you could perhaps infer a causal relationship" (id1) and another commented "you could conclude that one variable has an effect on another" (id8) and another simply stated "yes, one variable may cause another to happen" (id9). In contrast, just over half (55 per cent) of the students concluded no in response to this question because as one student wrote "correlational studies simply tell us that there is a relationship between two variables, not that one causes another" (id13).

Question 5. The data in Table 1 gives findings from a study of ten first year university students showing how much time they spent studying (on average per week through the year) and the end of year examination marks (out of 100). Plot the data on the graph (Figure 1) to make a scatter plot.

Table 1 Students' study time and examination marks

Student	study time	exam. mark
1	40	58
2	43	73
3	18	56
4	10	47
5	25	58
6	33	54
7	27	45
8	17	32
9	30	68
10	47	69

Figure 1 Scatter plot of study time by exam. performance



(Plots not provided to students in booklet)

None of the 20 students displayed any difficulties whatsoever in plotting the data set on the scatter plot. Evidently, students had the required skills for this question.

Question 6. What value does a perfect positive correlation coefficient take?

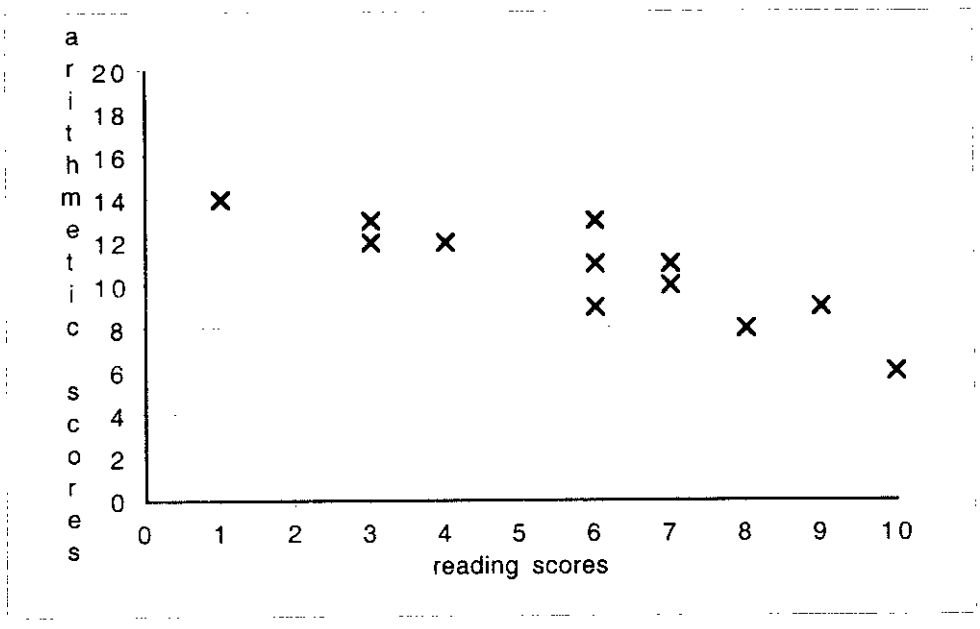
All but three of the twenty students wrote the value one in response to this question. Three of the students simply did not understand the question or did not know what the value of a perfect positive correlation coefficient was or, as one student wrote below the question were "?confused" (id12).

Question 7. In a study looking at the relationship between children's scores on reading test and their scores on an arithmetic test, the data shown in Table 2 was obtained. Plot the data on the graph. (Figure 2).

Table 2 Children's arithmetic and reading scores

Child	reading score	arithmetic scores
1	9	9
2	7	10
3	8	8
4	10	6
5	7	11
6	6	9
7	3	12
8	4	12
9	6	13
10	6	11
11	3	13
12	1	14

Figure 2 Scatter plot of reading scores by arithmetic scores



(Plots not provided to students in booklet)

Every single student correctly plotted this second scatter plot, but as the next question shows, they did not all interpret the graph in an appropriate way.

Question 8. What does the scatter plot show about the relationship between two sets of scores?

Although 50 per cent of the students did appropriately conclude that the scatter plot showed a negative correlation, 15 per cent did not pick on up this negative pattern that was shown between two variables. As one student wrote "there is not really a relationship between reading and arithmetic ability" (id16). Fifteen per cent of the students viewed the negative correlation displayed on the scatter plot as a positive correlation where, for example, one student answered "the scatter plot illustrates a positive correlation between arithmetic scores and reading scores" (id3) and another simply wrote "that there's a positive relationship" (id9).

Question 9. Which of the following are most likely to result in a high positive correlation and which are not likely to be correlated at all.

(i) Students' height and weight.

(ii) Girls' shoe size and scores on a reading test.

(iii) Students' self-ratings of ambitiousness and students' heights.

(iv) The number of theatre tickets sold and the number of customers in the audience.

Students did not display any difficulties in distinguishing between those pairs of variables that tend to result in a positive correlation and those that are not typically related because 80 per cent of the students indicated that students' height and weight, and the number of tickets sold and customers in the audience were likely to be positively correlated, and that girls' shoes size and scores on a reading test, and self-ratings of ambitiousness and height were not likely to be correlated. Three of the students, however, had an extra thought for the relationship between students' height and weight because they were not convinced that a high positive correlation would be found in this case. For example, one student

wrote "neither of these, because the taller you are has no bearing on your weight. Other factors determine how much you weigh not simply height" (id13).

Question 10. What is a likely correlation coefficient that you might obtain that would indicate no relationship between two variables. (For example, between students' self-ratings of ambitiousness and students' heights).

A student's answer to this question was categorised as appropriate if they wrote 0 (or if they provided a value very near 0). Twenty per cent of the students' answers to this question indicated that they either lacked the necessary knowledge to answer or were confused: "probably less than 3.8 or less than 2. something - this indicates no correlation" (id3). Fifteen per cent of the students thought that a perfect negative correlation would indicate no relationship between two variables. For instance, one of the students wrote "very close to -1. Probably -0.95" (id15).

Question 11. Which of the following five correlation coefficients represent the greatest amount of correlation? 0.5, -0.8, 0.2, -0.4, 0.

In response to this question, 60 per cent of the students indicated that -0.8 represented the greatest amount of correlation. And although one student could not answer the question at all and wrote "don't understand" (id2), 35 per cent of the students indicated that they thought a positive correlation represented the greatest amount of correlation in the set provided. In this case, six students chose 0.5 and one student chose 0.2 to represent the greatest amount of correlation.

Question 12. List the 5 correlation coefficients in order from those that indicate little or no correlation to that which indicates the greatest amount of correlation.

0.5, -0.8, 0.2, -0.4, 0.

In answering this question a student should indicate 0 for no correlation, -0.8 to represent the greatest amount of correlation, and rank the remaining coefficients irrespective of their

sign. One student clearly reasoned: "okay, no correlation, zero, I'm going on purely on the size ignoring whether it's positive or negative" (id5). (Excerpt from transcript). Twenty five per cent of the students appeared to view a positive coefficient as stronger than a negative coefficient. Students would therefore order or rank the coefficients as follows "0, -0.4, -0.8, 0.2, 0.5". To confuse matters further, 15 per cent of the students not only viewed a positive correlation as stronger than a negative one, but also thought that a negative correlation was indicative of no correlation and ordered the coefficients as follows "-0.8, -0.4, 0, 0.2, 0.5".

Question 13. For a study that was to look at the relationship between students' examination performance (measured by scores on a particular examination) and students' performance on course work (measured by marks for an assignment), state the null hypothesis.

This question did not elicit any misconceptions: students stated the null hypothesis correctly or they did not. The students did not appear to hold zany ideas about null. Indeed, 70 per cent of the students could state the null hypothesis as one of the student's answers exemplifies: "there will be no relationship between subjects' examination performance and their performance on course work" (id12).

Question 14. A psychologist was interested in the relationship between people's memory for shapes and their spelling ability, so she set up a study in which two tests were given to ten subjects. (Let's simply suppose that these two tests do in fact measure memory for shapes and spelling ability). The following Table 3 shows the scores that were obtained from the memory test for shapes and the test for spelling ability. State the null hypothesis and work out the value of the Pearson correlation coefficient, r. Use the formula for Pearson correlation coefficient provided.

Table 3 Subjects' memory test and spelling test scores

Subject	memory test	spelling test
1	6	3
2	7	4
3	8	2
4	9	4
5	4	1
6	3	1
7	9	5
8	8	5
9	10	6
10	11	5

Formula for Pearson correlation coefficient, *r*

$$r = \frac{N\sum a \times b - \sum a \times \sum b}{\sqrt{\left[N\sum a^2 - \left(\sum a\right)^2\right]\left[\left(N\sum b^2 - \left(\sum b\right)^2\right)\right]}}$$

where N = number of subjects

As the previous question showed, students can state the null hypothesis. In the first part of this question there were no surprises: 80 per cent of the students correctly stated the null hypothesis for the study.

This question also asked students to calculate the Pearson correlation coefficient by using the formula (and a calculator) that was provided. Only four out of the twenty students followed an appropriate procedure required by the formula to calculate *r*. Thirty

per cent of the students commented that they really did not understand what was required so that they could work out *r* and their responses were categorised as 'lacks knowledge'. Here, the space in the task booklet was often left blank. Students were prompted to give a reason or an explanation as to why they could not attempt a calculation of the coefficient. One student simply wrote "because I don't understand" (id12) and for one student the transcript for part of the session provides an explanation:

> S: "I've never done this before."
>
> E: "What have you never done before?"
>
> S: "I've never done a Pearson correlation."
>
> (id3). (Excerpt from audio recording of session).

It was evident that some students had not had to calculate a Pearson correlation coefficient by using a formula during their undergraduate studies. Fifty per cent of the students' answers to this question were categorised as 'problem with formula'. Students displayed a variety of symptoms that indicated that they had a problem with the formula, but primarily students could not derive a procedure to calculate *r*. One student was aware that she could not generate a procedure from the formula and wrote "can't use the Pearson correlation coefficient formula provided as it is not written out in a step by step pattern" (id8). More specifically, students had problems with the $\sum a \times b$ and/or $(\sum a^2)$ and/or $(\sum a)^2$ (or the equivalent $(\sum b^2)$ and $(\sum b)^2$.). Students were prompted as they tried to generate a procedure to calculate *r*, where, for example, it was suggested that they might create a additional column in the data table so that $(\sum a^2)$ could be calculated for the formula. So, students experienced problems in calculating the correlation coefficient because they found it difficult to generate a procedure from the formula, but students were also confused by the various symbols in the formula, such as the difference between $(\sum a^2)$ and $(\sum a)^2$. In addition, it was evident that some students did not expect that the calculated *r* should be between -1 and 1. For example, one student calculated *r* to be 59.27 (id3) and another student calculated it to equal 26686.372 (id14).

Question 15. Perform a two-tailed test to see if there is a significant relationship between the memory and spelling tests ($p < 0.05$). Here, use Table K provided.

The Table K that was provided to the students so that they could answer this question was a standard statistical table showing the critical values of *r* at various levels of probability. It also included the following information that is typically found on such tables: "the observed value of *r* is significant at a given level of significance if it is equal to or larger than the critical values shown in the table". Seventy five per cent of the students used the table correctly to see if the obtained value of *r* was significant. If the student could not work out the value of *r* in the previous question then they were provided with the correct value. One of the student's answers illustrates in detail the appropriate procedure that was followed by those students who used the statistical table to conclude that the obtained value of *r* was significant. This student wrote:

> "1) N - 2 = 8
>
> 2) Find 0.05 on 2-tailed test line
>
> 3) Go down columns to find result (.6319)
>
> 4) r is greater than this, so the result is significant"
>
> (id13).

Two of the students' responses were coded as 'lacks knowledge' because they could not answer question 15. For example, one of these students wrote "There is a significant relationship. (guessing)", but they did not use the statistical table. Three of the students' answers were placed in 'no category'.

Question 16. You have decided whether the calculated value of r is significant or not significant. What does your decision mean?

Seventy per cent of the students answered this question appropriately by suggesting that because the calculated value of *r* is significant this provided support to the idea that there is a relationship between people's memory for shapes and their spelling ability. For

example, one student wrote "there is a significant relationship between people's memory for shapes and their spelling ability" (id17). For this question, five of the students' answers were coded as 'no category'.

Question 17i. Part 1.

A real life concern that is often given media coverage is the effects of television on children's and teenagers' social behaviour. Does watching violence on television encourage aggression? Much media debate surrounds this issue, but it is an important one in this day and age: it has been estimated that the average child in the USA, by the age of sixteen, will have spent more time watching television than being in school, and will have seen 13,000 killings on television (Smith and Cowie, 1988). Psychologists have attempted to find out about the possible link that might exist between television violence and aggression. Let's take a study as an example. A group of researchers interviewed the parents of children who were 9 years of age (184 boys, 175 girls) to see what they favourite television programmes were. From this, the researchers constructed a measure of exposure to television violence. The children themselves were asked to rate the other children in their class for aggressiveness. The researchers found that the correlation between these two measures was 0.21 for boys, but only 0.02 for girls. As shown in Figure 3 provided, the correlation for the boys was significant ($p < 0.01$).

What are likely explanations for these findings?

Figure 3 Correlations between the amount of television violence viewed at age 9 and peer-rated aggression at age 9 (184 boys and 175 girls)

Boys

TV violence
at age 9

0.21**

aggression
at age 9

** $p < 0.01$

Girls

TV violence
at age 9

0.02

aggression
at age 9

(Adapted from Smith and Cowie, 1988, figure 4.4, p. 112. From Lefkowitz, Eron, Walder & Huesmann, 1977).

On the basis that students were provided with the findings of the study that were diagramatically presented in figure 3, a student's answer was categorised as appropriate if they said that the significant correlation found for the boys could mean either than viewing television violence caused aggression or that aggressive boys liked watching

violent television programmes. A student's answer was also categorised as appropriate if they suggested other plausible variables that could be responsible for the reported correlations in the figure. For example, parental discord in the home that could lead a child both to watch violent television programmes and also to be aggressive.

Only 20 per cent of the students provided an appropriate answer to question 17(i), where, for instance, one student commented "possibly the TV violence influenced aggression. Perhaps being aggressive creates a liking for more violent programmes" (id11) and another student concluded "for boys there is a correlation (positive) between parent's idea of children's favourite programmes and peer ratings" (id15). Thirty per cent of the students' answers could not be categorised because, for example, their answers were unclear. With the findings that were provided for the study concerning TV violence and aggression, four out of the twenty students inferred causality from correlation. One student commented "boys show statistical significance to being more aggressive as a result of watching TV violence more than girls. Tendency for boys to be more aggressive than girls due to watching violent TV programmes" (id1) or, as another student put it, "boys: positive correlation that is higher than the girls. Therefore boys exposed to more aggression and therefore display more aggression whereas girls are not exposed to as much and therefore don't display as much" (id18). These students do not consider or suggest other possible mediating variables between TV violence and aggression at age nine.

A category called 'comparison of sex differences' was created to account for 25 per cent of the students' answers to this question. Here, students did not focus on the significant correlation found between TV violence and aggression at age 9 or the non significant relationship that was reported for the girls, but instead they talked about boys being more aggressive and/or boys watching more violent TV than the girls. With regard to the reported findings, this is not necessarily the case. One student made a comparison in this vein where they concluded "that boys tend to watch more television than girls and

that they are rated as being more aggressive than girls in the class" (id3) and another student wrote "an explanation could be that the boys watched more violence on TV than girls" (id8). And although one student made a valuable point about one of the methods used in the study, they still concluded that more boys were viewed as aggressive: "I feel that both boys and girls are likely to rate boys as more aggressive hence the higher amount of boys perceived to be aggressive" (id17).

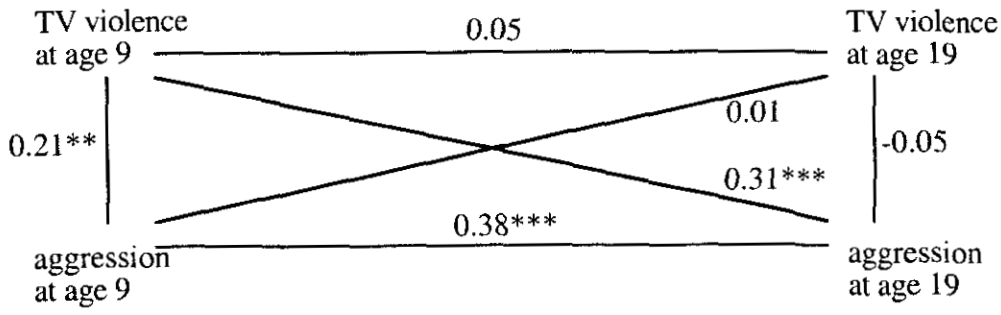
Question 17ii. Part 2.

Ten years later when the children were teenagers (19 years old), the same measures were taken. The correlations between the same two measures at this time and the correlations between the two different time periods are shown for both males and females in Figure 4 provided.

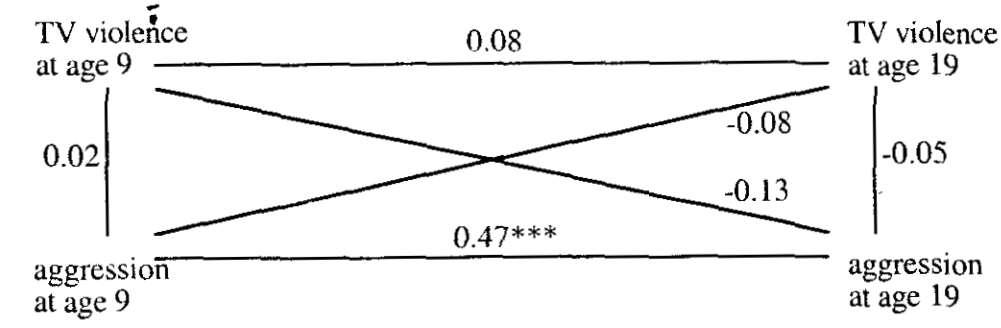
What do these findings suggest?

Figure 4 Correlations between the amount of television violence viewed at ages 9 and 19 and peer-rated aggression at ages 9 and 19 (184 boys and 175 girls)

BOYS



GIRLS



**p<0.01 ***p<0.001

(Adapted from Smith and Cowie, 1988, figure 4.4, p. 112. From Lefkowitz et al, 1977).

Here, students' answers were categorised as appropriate if, for example, they picked up on the significant correlations and/or the cross-lagged correlations that were displayed in figure 4. One student used the figure to summarise the pertinent findings as follows:

"1. Significant correlation between boys aggression at 9 and at 19.

2. Significant correlation between boys TV violence at 9 and aggression at 19.

3. Significant correlation between boys aggression at 9 and TV violence at age 9.

4. Significant correlation between girls aggression at 9 and aggression at 19.

5. Negative correlation between girls aggression at age 9 and TV violence watched at age 19"

(id14).

Although five out of the twenty students did interpret the figure appropriately to summarise the study's findings, they did not provide possible reasons for the links that were reported between the variables. Forty five per cent of the students did not successfully interpret the figure because they did not identify the significant correlations and, for example, contrast these with the non significant relationships that were shown on the figure. Of these students, one only commented on one aspect of the figure and wrote that they "looked at correlation between boys TV violence at 9 and aggression at age 19. Weak positive correlation, so can't say that watching TV violence makes you more aggressive" (id13). Fifteen per cent of the students made it clear in the task booklet that they did not understand the figure: one student wrote "can't understand the table" (id8) and another stated "don't understand" (id12).

3.4.4 The piloting of tasks

One of the objectives of the study was to investigate the kinds of questions that can be used to assess students' understanding of correlation. It was evident that some questions were more valuable and useful in this respect than others. Questions 1, 4, 8, 10 and 12 uncovered and made clear particular conceptions that students held.

A student's answer was placed in 'no category' if, for example, their answer was unclear, or they failed to answer the question. Even with careful prompting some of the students' answers reflected the fact that they were often very confused by the question or evidently lacked the required knowledge. Some students exhibited a unique conception. A category was created if three or more of the students held a particular misconception in common, but in some cases only one or two students answered a question along similar lines. If this was the case, their answer(s) were also coded as 'no category'. It was found that questions 3 and 17(i) resulted in a high frequency of answers that were difficult to categorise (e.g., 25 per cent of the students' answers to question 3 were coded as 'no category').

Only two categories were required to account for all of the students' answers to question 6. Students either knew that a perfect positive correlation coefficient takes the value of one or they did not. Similarly, only two categories ('normal answer' and 'lacks knowledge') were necessary to account for the students' responses to question 11 where students were asked to choose a correlation coefficient that represented the greatest amount of correlation. The students' answers to this question did not provide reasons why, for example, they chose 0.5 rather than -0.8 to represent the greatest amount of correlation, but it is a valuable question because it confirms the findings from question 12 that indicated that seven out of the twenty students had difficulty with the concept of correlation.

Question 14 asked the students to work out by hand a correlation coefficient. In the main, students found this difficult because they did not have a step by step procedure to hand that they could follow. In addition, as part of their undergraduate studies, students at one of the universities had not been required to work out this particular statistic by hand. In answering question 14, students tended to be prompted more than they were on any of the other questions. The prompts helped them to decide what they could do next as

they tackled the formula and to clarify any confusions that they might have had about particular symbols, (such as the meaning of Σ).

Question 9, which asked students to decide which pairs of variables were positively correlated or not, did not invoke any misconceptions concerning correlation that students might hold. Questions 13, 15 and 16 simply showed that the majority of students could state the null hypothesis for a study, see if a calculated value of a correlation coefficient was significant, and decide that this meant that there was a significant correlation between two variables.

3.5 Discussion

3.5.1 Statistics as difficult

In this study 60 per cent of the students described statistics as a difficult subject to study. The task booklet that the students completed in the study provided an indication of a student's understanding of correlation and the associated skills in this area, but the overall score for this booklet was not related to whether students described statistics as difficult or not.

3.5.2 Tasks to assess students' understanding of correlation

In terms of assessing students' understanding of correlation, some of the devised questions were useful because they invoked students' misconceptions. Some questions, however, generated student responses that were not easily categorised.

The wording of question 11 and question 12 was ambiguous because they asked the student which correlations indicated "the greatest amount of correlation". A student might have equated the word greatest with the relative value of a whole number. Versions of these questions that were used for subsequent research were modified to ask the student which correlation coefficients indicated "the strongest" correlation or relationship.

Subsequent research concerning students' conceptions in correlation, which is described in this thesis, used modified versions of the following questions in the task booklet: question 4, question 8, question 10, question 11 and question 12. A wider selection of texts were used for this further research to develop additional questions that were designed to assess students' understanding of correlation.

3.5.3 Students' difficulties with statistical procedures

A majority of the students that took part in the study experienced difficulties in deriving a statistical procedure from a formula so that they could work out a statistic. This finding indicates that students lacked the prerequisite mathematical skills for statistical techniques. At one of the universities, one of the lecturers commented that the students who were likely to take part in the study would not have been required, as part of their undergraduate studies, to work out Pearson correlation coefficient by hand. It is, however, suspected that students were used to having a step by step procedure or a worked example to calculate a particular statistic (see Greene & d'Oliveira, 1982), and that they lacked the required mathematical skills to use a formula. To generate the appropriate procedure from a statistical formula a student must have acquired prerequisite mathematical skills, such as how to correctly perform operations that involve parentheses. (See Minium, 1978, pp. 471 - 487).

Many of the students who took part in the study found it difficult to interpret data that was provided in the context of a study in psychology. Given a set of data that was diagramatically presented, students did not tend to focus on those correlation coefficients that were significant or suggest the possible reasons for the relationships.

In the previous chapter it was asked whether psychology students lack prerequisite mathematical skills for statistical techniques or hold statistical misconceptions that impede the acquisition of statistical concepts. The findings of this investigation indicated that

students lacked the required skills to use statistical techniques involving formula, but it was also found that they held statistical misconceptions.

3.5.4 Students' misconceptions about correlation

The findings indicate that some students have confusions about correlation. Not all of the students had a clear idea of a correlational design. In addition, some students inferred causality from correlation. A number of the questions indicated that students had difficulties with the concepts of positive correlation, negative correlation and no correlation: some assumed that a positive correlation is stronger than a negative correlation and/or that a negative correlation is indicative of no correlation between two variables.

These findings that concern students' misconceptions in correlation needed to be confirmed by further research that involved a different population of students who attended other institutes of higher education. This was necessary because particular approaches to instruction, which may lead students to acquire particular misconceptions, might be practised at one institution but not at another. Inevitably students at one university will be taught about correlation in a particular way and this may affect their understanding of the area. Chapter 8 describes empirical work that was conducted concerning students' misconceptions that involved students who attended a residential school for a psychology course. The findings of this study confirmed the findings of the investigation described in this chapter: students have confusions about causality, negative correlations and the strength of correlations.

As noted in chapter 2, one recent research programme has looked at pre-university students' understanding of association and correlation (Batanero et al, 1996; Batanero et al, 1997; Estepa & Batanero, 1996). Batanero and her colleagues (1997) have been concerned with students' understanding of association in general and have conducted empirical work that has looked at students' conceptions about association in contingency

tables (Batanero et al, 1996) and at students' conceptions about correlation in scatter plots (Estepa & Batanero, 1996). The research study described in this chapter was conducted in 1996 and so the findings of Batanero et al's (1996, 1997) work did not inform the design of this study. Batanero and her colleagues' work has used a different population of students from the study that is described in this chapter. The investigation described in this chapter has looked specifically at university psychology students' understanding of correlation and their skills in this area.

Estepa and Batanero (1996) investigated pre-university students' strategies when they assessed correlation in scatter plots. A sample of 213 students who were in the last year of secondary school (18 years old) and who had not received instruction concerning statistical association were recruited for this investigation. The primary aim of the research was to identify students' preconceptions about correlation, and the students' judgments and strategies in their assessment of correlation in scatter plots were used as indicators of such conceptions. Students completed a questionnaire that contained three questions concerning scatter plots. For example, item three on the questionnaire provided a scatter plot that represented a correlation of 0.55 and asked:

> "Two judges ... qualified twenty students who took part in a project competition in accordance with their opinion. We have represented in this plot the place that each student was given in both judges' classification. We would like to study if there is some relationship between the place assigned to the same participant by the two judges (if the judgment of both judges is or not related). ... Do you believe that the relationship between the place assigned to the same student by both judges is direct, inverse or that there is no relationship at all? (Explain your answer)" (Estepa & Batanero, 1996, p. 41).

In contrast, item one provided a scatter plot that showed no relationship between two variables (independence) and represented a correlation of 0.11, and item two provided a scatter plot that showed a negative correlation (inverse) of -0.77 between two variables.

Task variables such as the sign and value of the correlation were considered on the three items of the questionnaire. Students' answers to each of the items were analysed to see which type of correlation was perceived by the students (direct, inverse or independence). In addition, students' answers were categorised in terms of whether they employed correct, partially correct or incorrect strategies in their judgment of correlation in the scatter plots.

Generally, students were able to judge the type of correlation between the variables. For example, for item one (no correlation) 83 per cent of the students provided the correct relationship, (although 11 per cent judged that the relationship was negative.) Similarly, a majority of the students (85 per cent) judged appropriately that item two showed a negative correlation. Five per cent of students, however, thought it was a positive relationship and nine per cent (8.9) thought that the scatter plot for item two showed no relationship. These are similar findings to the ones obtained in the investigation described in this chapter: 50 per cent of the participants appropriately concluded that the scatter plot in question 8 showed a negative correlation, but 15 per cent of the participants simply did not pick up on the relationship and 15 per cent viewed the negative correlation displayed on the scatter plot as a positive correlation.

Some students did, however, experience difficulty with item three concerning the relationship of the judgment of two judges that is detailed above. Here, only 22 per cent of the students provided the correct relationship (positive correlation) and 64 per cent of students decided that there was no relationship between the variables. This low percentage of correct answers for this item might be due to the spread of the scatter plot and to the fact that the relationship between the two variables was not due to any causal influence (Estepa & Batanero, 1996).

Students' answers were also analysed with respect to strategies employed in the assessment of the relationship in each item. These strategies served to identify incorrect conceptions or misconceptions (Estepa & Batanero, 1996). These researchers have noted

that the idea of correlation is not simple and that students may have correct intuitive conceptions about properties related to correlation as well as some incorrect conceptions which might lead them to adopt an incorrect or partially correct strategy. One incorrect strategy was described as follows:

> Causality. In spite of an empirical correlation the student argued that there was no correlation because one variable could not cause a direct influence on the other variable. This answer indicated a causalistic conception of correlation.
> (Adapted from Estepa & Batanero, 1996, p. 33).

In the case of item three, 24 per cent of the students used the causality strategy (Estepa & Batanero, 1996). In the empirical study described in this chapter, a causalistic conception of correlation was identified if a student simply inferred causality from correlation. Estepa and Batanero (1996), however, have taken a rather different angle in describing a causalistic conception. If a student explained that there was no correlation between two variables because one of the variables could not cause a direct influence on the other variable, then this was considered to be an indicator of a causalistic conception of correlation (Estepa & Batanero, 1996).

From empirical research that has investigated students' judgments of association in contingency tables and scatter plots the following conception has also been identified (Batanero et al, 1997; Estepa and Batanero, 1996):

> "Unidirectional conception of association. Sometimes students perceive the dependence only when the sign is direct, and so they consider an inverse association as independence" (Batanero et al, 1996, p. 166).

It is not clear, however, if this conception was evident when students judged association from contingency tables and/or from scatter plots (Batanero et al, 1997). Batanero et al (1997) do, however, provide a case study of the learning process of a pair of students in a computer-based teaching environment that covered the topics of association and correlation. In one of the teaching sessions, the pair of students was observed to be

surprised when they met a negative correlation coefficient for the first time and asked their teacher if this was possible. It was also evident that the knowledge of the properties of negative number ordering impeded the students' learning about negative correlation (Batanero et al, 1997). The findings of the investigation described in this chapter also indicated that students had difficulties concerning negative correlations. For example, 25 per cent of students appeared to think that a positive correlation coefficient was stronger than a negative coefficient when this was not the case.

3.6 Summary

The research study described in this chapter investigated students' subjective opinions about studying statistics: it was found that 60 per cent of the students in the study described statistics as difficult. A task booklet was used in the study to ascertain students' difficulties pertaining to the topic of correlation. It was found that students had difficulties in deriving a statistical procedure from a formula to obtain a statistic, in interpreting correlations and had a variety of conceptual confusions.

Five questions that concerned causality, negative correlations and the strength of correlations, which were used in the research study, identified particular misconceptions in correlation that students hold. These questions were later modified and used for further research described in this thesis.

The investigation found that students inferred causality from correlation. The term causalistic conception, which is used by Batanero and her colleagues (1996), will be subsequently used in this thesis to describe the idea that given a correlation, a student states that one variable has a direct causal influence on another variable and he/she does not entertain any other possibilities by, for example, suggesting that a third variable could be responsible for the obtained correlation.

It was found that students thought that a negative correlation coefficient was indicative of no correlation. Students also thought that a negative correlation that was

displayed on a scatter plot indicated no relationship between variables or that there was simply a relationship between the variables, but they did not specify the type of relationship. These confusions describe a unidirectional conception of correlation. The term unidirectional conception of correlation was used by Batanero et al (1996) to describe the following finding: "students perceive the dependence only when the sign is direct, and so they consider an inverse association as independence" (p. 166). In other words, students tended to only see a correlation when it was positive and thought that a negative correlation indicated no relationship between variables. These findings are very similar to those of the investigation described in this chapter and accordingly, a unidirectional conception of correlation is adapted as a term and used to describe one or more of the following:

- A negative correlation coefficient is thought to indicate no correlation.
- A negative correlation that is displayed on a scatter plot is viewed simply as a relationship or as a positive relationship.
- A negative correlation that is displayed on a scatter plot is described as indicating no relationship between two variables.

The findings of the investigation suggested that students were confused by the strength of correlations: students indicated that a positive correlation coefficient was stronger than a negative correlation coefficient when this was not the case. In addition, students indicated that in this case either a negative correlation coefficient was stronger than no correlation or that a negative correlation indicated no correlation between variables.

In chapter 2, it was argued that students' understanding of particular areas in statistics must be investigated so that misconceptions can be identified. This means that appropriate instructional materials can be designed to address these difficulties. In the following chapter, it is emphasised that learning is cumulative and that students' prior

knowledge in the form of common confusions relating to particular concepts must be addressed in the design of computer-assisted learning programs. To reiterate, computer-assisted learning programs are likely to be used increasingly as part of the statistics curriculum and could therefore provide an additional form of instruction to help students learn about statistical concepts. Indeed, the following chapter reviews research and developments in the field of computer-assisted learning for statistics and examines existing programs for correlation.

Chapter 4

Computer-assisted learning for statistics

4.1 Introduction

This chapter looks at research-based principles of learning and research and developments in the field of computer-assisted learning for statistics. The systems *Stat Lady* (Shute et al, 1996) and *StatPlay* (Cumming & Thomason, 1998) are critically examined in this chapter. These systems are examined because they are the main computer-assisted learning programs for statistics that have associated research programmes, which are reported in the research literature.

There are a number of computer-assisted learning programs that include sections that are designed to teach correlation to students in higher education. The design of *Link* was initially informed by research literature relating to computer-assisted learning for statistics, and an examination of sections of computer-assisted learning programs that cover correlation. A critical review of relevant parts of these programs is provided in this chapter.

4.2 Learning statistics constructively

One of the reasons why psychology students might find statistical concepts difficult to acquire is that they are presented with concepts that are not related to psychological research. The acquisition of statistical concepts is likely to be facilitated if such concepts are illustrated with reference to an interesting psychological study. Indeed, psychology students must not only learn which statistical techniques should be used to analyse data that is collected from their experiments or projects, but they must also learn to interpret the statistics obtained in relation to empirical questions and statistics that are presented in psychology journals and research papers.

Constructivist conceptions of learning have described student learning as active, constructive and cumulative (Shuell, 1992). These conceptions of learning have highlighted the importance of prior knowledge, learner engagement, learner activity and the context of learning as students acquire concepts and skills in subject areas (e.g., Bransford et al, 1990; Brown, Collins & Duguid, 1989; Choi & Hannafin, 1995; Cognition and Technology Group, 1992). Learning is cumulative in that new learning is built on and influenced by a learner's prior knowledge (De Corte, 1995; Shuell, 1992). It has been argued that not only should the learner be active in the construction of knowledge, but that the subject matter to be learnt must be situated in meaningful and realistic environments of contexts (Cognition and Technology Group, 1992). Indeed, the design of the program *Stat Lady*, which is reviewed in this chapter, was informed by

> "theoretical postulates that learning is a constructive process, enhanced by experiential involvement with the subject matter that is situated in real-world examples and problems" (Shute et al, 1996, p. 25).

The concept of anchored instruction was proposed in response to apparent failures in traditional education where students typically treat new knowledge as facts and procedures to be learnt rather than knowledge to be used (Bransford et al, 1990). As a pedagogical approach, however, anchored instruction has only provided generic prescriptions for instruction and has been ambitious in its application:

> "anchored instruction, represents an attempt to help students become *actively engaged in learning by situating or anchoring instruction in interesting and realistic problem-solving environments.* These environments are designed to invite the kinds of thinking that help students develop general skills ... plus acquire specific concepts and principles that allow them to think effectively about particular domains" (emphasis added. Cognition and Technology Group, 1992, p. 135).

This group of researchers have proposed that the work on anchored instruction was designed to be relevant to instruction in all content areas from reading and writing to mathematics. For present theoretical purposes, an important tenet of anchored instruction is that instruction is situated or anchored in an interesting and meaningful context. Savery and Duffy (1995) have proposed a similar instructional principle that was derived from constructivism: "anchor all learning activities to a larger task or problem" (p. 32). The learning environment, task or problem in which learning activities are situated will inevitably vary depending on the subject matter to be learnt.

It is therefore proposed that the acquisition of statistical concepts is facilitated if they are presented with reference to interesting psychological research. A study in psychology should be used to provide a context to present and illustrate statistical concepts. However, if statistical concepts are to be successfully acquired, then an instructional approach must also address psychology students' prior knowledge.

4.3 *Stat Lady*

One research programme has involved the development and evaluation of a computer-assisted learning program called *Stat Lady* (Shute et al, 1996; Shute & Gawlick-Grendell, 1994; Shute & Gawlick-Grendell, 1993). To date, *Stat Lady* consists of two modules that cover the topics probability and descriptive statistics, but only the description of the probability module and its evaluation have been reported in the literature (Shute et al, 1996; Shute & Gawlick-Grendell, 1994). For this thesis, the *Stat Lady* program was not available for research review purposes and screen shots of the program could not therefore be provided. Shute and Gawlick-Grendell (1993) have reported that the design of *Stat Lady* was based on models of learning (e.g., Cognition and Technology Group, 1992) which have emphasised that learning is facilitated with student involvement and is enhanced or even optimised if concepts, principles and procedures are situated or anchored in real world scenarios and problems.

The primary aim of *Stat Lady* is to aid the acquisition of statistical knowledge and skills by making learning meaningful and therefore memorable. To do this, *Stat Lady* provides activities, such as painting cars, rolling dice and betting that were devised to be entertaining and related to every day examples (Shute et al, 1996). For example, *Stat Lady* provides a betting game that was designed to engage the student in the learning process. Here, learners are provided with $5-00 electronic cash and bets are submitted on different combinations of numbers. A particular game can then be defined where, for example, the learner wins the bet if an 11 or 12 appears on a roll of two dice and *Stat Lady* wins if 9 or 10 appears on a roll. The student makes a bet by selecting buttons at the human-computer interface which causes *Stat Lady*, (a female figure with glasses on the screen), to shake and roll two dice. Learners come to realise that over time and through the loss of electronic dollars, most of the games are unfair to the learner. To show this they must construct a table that lists all two dice events and all possible outcomes corresponding to each event and associated probabilities. Learners can then assess the 'fairness' of games by working out the probability of obtaining a 9 or 10 versus an 11 or 12 (Shute & Gawlick-Grendell, 1993).

Stat Lady also provides a mastery learning approach to instruction where a concept or principle is presented and then a problem is presented for the student to solve so that they may demonstrate comprehension of part of the curriculum. Although Shute et al (1996) have stated that learning with *Stat Lady* is self-paced, students must solve at least two problems correctly before they can move on to another part of the curriculum. If a student's answer to a problem is incorrect, specific feedback is provided and the student can attempt to answer the question again. *Stat Lady* contains a buggy library of known and likely incorrect responses to the statistical problems. For each of the problems in the system, a range of conceivable mistakes that students might make in answering particular questions was predetermined by pilot testing of the system and consultation with statistics experts (Shute et al, 1996).

The primary aim of *Stat Lady* is to aid the acquisition of statistical knowledge and skills by making learning meaningful and therefore memorable. To do this, *Stat Lady* provides activities, such as painting cars, rolling dice and betting that were devised to be entertaining and related to every day examples (Shute et al, 1996). For example, *Stat Lady* provides a betting game that was designed to engage the student in the learning process. Here, learners are provided with $5-00 electronic cash and bets are submitted on different combinations of numbers. A particular game can then be defined where, for example, the learner wins the bet if an 11 or 12 appears on a roll of two dice and *Stat Lady* wins if 9 or 10 appears on a roll. The student makes a bet by selecting buttons at the human-computer interface which causes *Stat Lady*, (a female figure with glasses on the screen), to shake and roll two dice. Learners come to realise that over time and through the loss of electronic dollars, most of the games are unfair to the learner. To show this they must construct a table that lists all two dice events and all possible outcomes corresponding to each event and associated probabilities. Learners can then assess the 'fairness' of games by working out the probability of obtaining a 9 or 10 versus an 11 or 12 (Shute & Gawlick-Grendell, 1993).

Stat Lady also provides a mastery learning approach to instruction where a concept or principle is presented and then a problem is presented for the student to solve so that they may demonstrate comprehension of part of the curriculum. Although Shute et al (1996) have stated that learning with *Stat Lady* is self-paced, students must solve at least two problems correctly before they can move on to another part of the curriculum. If a student's answer to a problem is incorrect, specific feedback is provided and the student can attempt to answer the question again. *Stat Lady* contains a buggy library of known and likely incorrect responses to the statistical problems. For each of the problems in the system, a range of conceivable mistakes that students might make in answering particular questions was predetermined by pilot testing of the system and consultation with statistics experts (Shute et al, 1996).

A target audience of students was not explicitly considered in the design of *Stat Lady*. Is the system to be used by students in higher education who are training to be statisticians or by students studying economics or psychology? Related to this is the approach that was taken in the design of the curriculum that *Stat Lady* was designed to impart:

> "*Stat Lady's* knowledge base was initially developed through careful inspection of six introductory statistics textbooks ... similar to the instructional content and sequencing within the majority of textbooks, *Stat Lady's* curriculum began with the explication of simple concepts ... then progressed to the instruction of various probability rules" (Shute et al, 1996, p. 28).

It could be argued that a computer-assisted learning program should not simply include materials that were covered in some form in a number of textbooks. Rather, the design and development of a learning program should also address what topics are typically covered by the target audience of the system and investigate those areas that students find difficult. A developed program can offer an alternative form of instruction for tricky topic areas and, for example, be specifically designed to address statistical misconceptions. *Stat Lady's* buggy library was developed from pilot testing the system and consultation with statistics experts so that a range of conceivable mistakes that students make in solving different problems could be determined. However, the comprehensive literature on people's statistical misconceptions (e.g., Garfield & Ahlgren 1988; Shaughnessy, 1992) has not been recognised in Shute and her colleagues' research programme.

An additional concern is the inappropriate description of *Stat Lady* as "being an expert statistician" and "an avid story-teller" (Shute et al, 1996, p. 28), and "unlike some Statistics professors, *Stat Lady* is consistently good-natured ... talking with learners, not at them" (Shute & Gawlick-Grendell, 1994, p. 180). Although comprehensive evaluation studies of the program have been conducted, the instructional capabilities and possibilities of *Stat Lady* have been over-stated:

"a very experiential learning environment with ... enticing displays ... thus *empowering* ... learners rather than simply providing formulas to memorise or tables of numbers to manipulate" (emphasis added. Shute et al, 1996, p. 28).

Two evaluation studies of *Stat Lady* have been conducted (Shute et al, 1996; Shute & Gawlick-Grendell, 1994). The most recent evaluation study, which was conducted to test the efficacy of learning probability from *Stat Lady* as compared to a traditional lecture (Shute et al, 1996), is briefly discussed here. The participants of this study, who signed up to take part in the *Stat Lady* group, the lecture group or the control group, were undergraduate psychology students at an American university. All of the participants completed a pre test at the beginning of the week and then received three hours of instruction on three consecutive days, in one hour periods of instruction from either *Stat Lady* or a lecture. At the end of the week the students completed a post test, and the participants in the no-treatment control group, who did not use *Stat Lady* or attend the lectures, simply completed the post test at the end of the week. It was found that both the *Stat Lady* and the lecture groups showed a significant increase from the pre test to the post test, but the control group did not. However, in terms of students' learning outcomes as measured by the pre and post tests, the *Stat Lady* group and the lecture group did not differ significantly from each other.

Similar findings to that of the above were found in a study that looked at students learning probability from *Stat Lady* or from a paper and pencil workbook version of the probability curriculum implemented in *Stat Lady*: performance on the learning outcome measures did not differ significantly between the two experimental conditions, but both groups performed significantly better than the control group (Shute & Gawlick-Grendell, 1994).

It is not clear from these evaluation studies which specific aspects of *Stat Lady's* design contributed to students' learning gains from the pre to post test. The above studies

will be briefly re-considered when methodology for the evaluation of computer-assisted learning programs is discussed in the next chapter.

Despite the above criticisms, Shute and her associates (1996) have contributed to research by designing a system that is based on the theoretical postulates that learning is a constructive process and is facilitated if subject matter is anchored to real world examples. In addition, comprehensive evaluation studies of a system that was designed to teach probability have been conducted.

4.4 *StatPlay*

Cumming and his associates have referred to people's naive statistics that impede the acquisition of statistical concepts. They have stressed that naive statistics, in the form of common statistical misconceptions, are resistant to traditional forms of instruction (Cumming & Thomason, 1998; Cumming & Thomason, 1995; Thomason, Cumming & Zangari, 1994). These researchers have proposed that simulations, demonstrations and multiple representations that are dynamically linked, which have been implemented in a program called *StatPlay*, can help students overcome their naive statistics and provide a promising approach "to promote true conceptual change" (Thomason et al, 1994, p. 59). At the Fifth International Conference on Teaching of Statistics (Singapore, 1998), Cumming stated that it was planned that *StatPlay* would be available for teaching purposes in 1999.

StatPlay was designed to be used on introductory statistics courses in any discipline (Cumming & Thomason, 1998). One of the primary goals of *StatPlay* was to help students overcome statistical misconceptions about central statistical concepts. Drawing on research (e.g., Tversky & Kahneman, 1993), Cumming and Thomason (1995) pointed out the following prevalent statistical misconceptions where people generally:

(i) Underestimate the amount of variability from sample to sample if repeated samples are taken.

(ii) Overestimate the extent to which a sample is likely to reflect the properties of the population from which it is drawn.

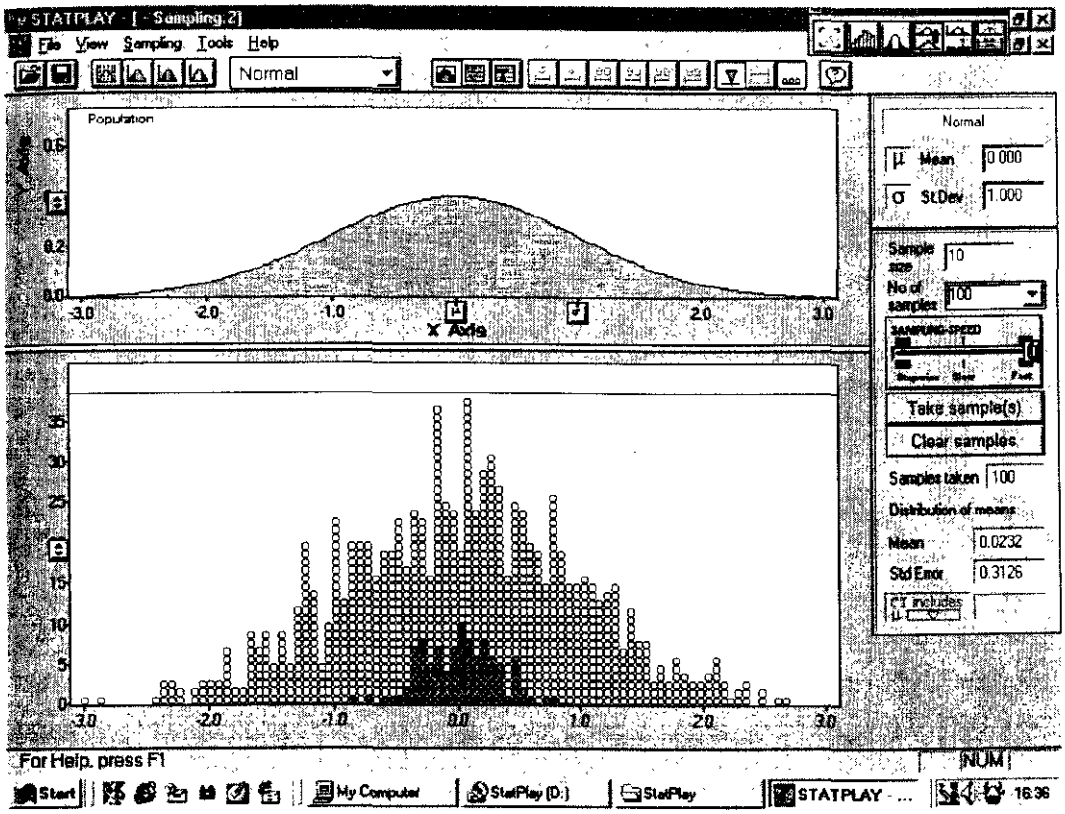
(iii) Expect that the repetition of an experiment is much more likely to give a similar outcome than is in fact the case.

(iv) Do not realise the important role of sample size, N in (i), (ii) and (iii) above.

As shall be shown below, it was necessary to address the idea of sampling variability in the design of *StatPlay*.

To illustrate *StatPlay*, the *Sampling Playground* is briefly described here. In this playground students can explore variability in sample means and can choose to sample repeatedly from a population of any shape by taking samples of any size that they specify (figure 4.1). This playground provides a graphical display of the population, values of the population parameters and controls, so that the size and number of samples to be successively taken from the population can be specified by the user. The results of sampling from a specified population can be viewed graphically where the sample means are shown in relation to the population mean. The idea that there is variability in sample means can be introduced to students by using the *Sampling Playground* (Finch & Cumming, 1998).

Figure 4.1 The Sampling Playground in StatPlay



StatPlay provides two design features that are thought to assist understanding of statistical concepts: dynamically linked multiple representations and "vivid take-home images" (Cumming & Thomason, 1998, p. 948). Cumming and Thomason (1998) have argued that if a person can describe a concept in words, provide a definition, sketch an appropriate picture, write an associated formula, explain an application of the conception and realise how such different representations relate, then they have conceptual understanding. They have proposed that

> "Working with multiple, linked representations should help students develop such rich understanding" (ibid., p. 948).

StatPlay should help students overcome their statistical misconceptions because they can use multiple representations that are computationally linked (Cumming & Thomason, 1998). *StatPlay* consists of a number of playgrounds and in the *Distributions Playground* of the program, a curve of a distribution and the numerical values of the distribution

(mean and standard deviation of the distribution) are shown simultaneously and linked dynamically so that if, for example, the mean and standard deviation of the distribution are changed by the learner they will see corresponding changes in the shape of the distribution (Cumming & Thomason, 1995).

It has also been proposed that the representations of *StatPlay* are so vivid that they become "take-home images" for students (Cumming & Thomason, 1998). It is not clear what these researchers mean by these images. It is assumed that the representations provided in *StatPlay* illustrate statistical concepts visually and clearly and might therefore make an impression on the learner.

StatPlay has undergone prototyping and to date the development of a new version of the program is in progress which will involve the addition of more playgrounds including one on correlation and regression (Cumming & Thomason, 1998). A variety of evaluations studies have been conducted involving psychology undergraduates using *StatPlay* (Finch & Cumming, 1998). It is difficult to judge the impact of *StatPlay* on students learning statistics because details concerning the number of students that were involved in some of these evaluation studies have not been provided and students' understanding of relevant statistical concepts was not always assessed prior to using *StatPlay*. Two evaluation studies that did involve assessing students' understanding of relevant statistics prior to using *StatPlay* and after its use are briefly considered.

In these two evaluation studies, Finch and Cumming (1998) looked at students using *StatPlay* in relation to two instructional goals: (a) to acquire the idea that means vary from sample to sample and (b) to acquire sensitivity to the changes in this variability with respect to sample size. One study employed a pre-test-post-test quasi-experimental design, where psychology students saw a lecture demonstration, (but it is not clear whether *StatPlay* was used in this lecture demonstration), and then carried out activities using *StatPlay*. Students estimated variability in means for different sample sizes prior to using *StatPlay* and students showed almost no sensitivity to changes in sample variability

with changes in sample size (mean sensitivity score = 1.1, where 1 indicated no sensitivity). Having used *StatPlay*, students provided accurate estimates of relative changes in variability with sample size (mean = 1.6, where 1.8 indicated appropriate sensitivity) (Finch & Cumming, 1998). Finch and Cumming (1998) have not reported the number of students who took part in this study and the pre and post test change was not reported to be significant. Moreover, it is not clear whether the lecture demonstration or the actual student use of *StatPlay* contributed to students' learning outcomes. In addition, details of how the sensitivity scores were assessed have not been provided.

An experimental study was conducted which looked at students who worked with *StatPlay* activities and students who worked on similar activities but with pencil, paper and calculators. A substantial increase in students' sensitivity to sample size was observed in the *StatPlay* group (from 1.4 to 1.9 on the sensitivity scale) and no change was observed in the control group (Finch & Cumming, 1998). However, these findings were not reported as statistically significant and the number of students who participated in this study was not reported.

Finch and Cumming (1998) have pointed out that the above two studies have indicated the potential of *StatPlay* in "promoting conceptual change" (p. 899). Clearly, the evaluation of *StatPlay* requires further empirical work and studies must be designed to specifically look at, for example, the use of dynamically linked multiple representations in addressing students' statistical misconceptions.

4.5 Computer-assisted learning programs for correlation

Prior to the design of *Link*, it was recognised that it was imperative that existing computer-assisted learning programs, which cover correlation, must be examined and reviewed. The findings of the study reported in chapter 3 showed that students tend to hold misconceptions relating to correlation. Specifically, students have misconceptions concerning negative correlations and the strength of correlations, and infer causality from

correlation. The review process was therefore informed by the findings of the investigation described in the previous chapter and the computer-assisted learning programs were assessed with regard to these findings.

This process of review had three main objectives. Firstly, it was necessary to avoid the re-invention of design features of an existing program. Secondly, it was thought that a comprehensive review of available programs would indicate the kinds of learner activities that could be used in a computer-assisted learning program. Thirdly, the review would highlight limitations of existing programs.

This review focused on the sections of computer-assisted learning programs that cover correlation, but a related project has involved the formal evaluation of computer-based teaching resources for statistics, which included web-based materials (Morris & Le Voi, 1998). The review described here involved the critical examination of sections of computer-assisted learning programs, which were inspected if they covered correlation and were targeted at psychology or social science students in higher education. Since this review process, which informed the design of *Link*, further titles, such as *ActivStats* (*ActivStats*, 1997) have been released. Such recent developments could not obviously be reviewed prior to the development of *Link*, and they are not therefore considered in this thesis.

4.5.1 *ConStatS*

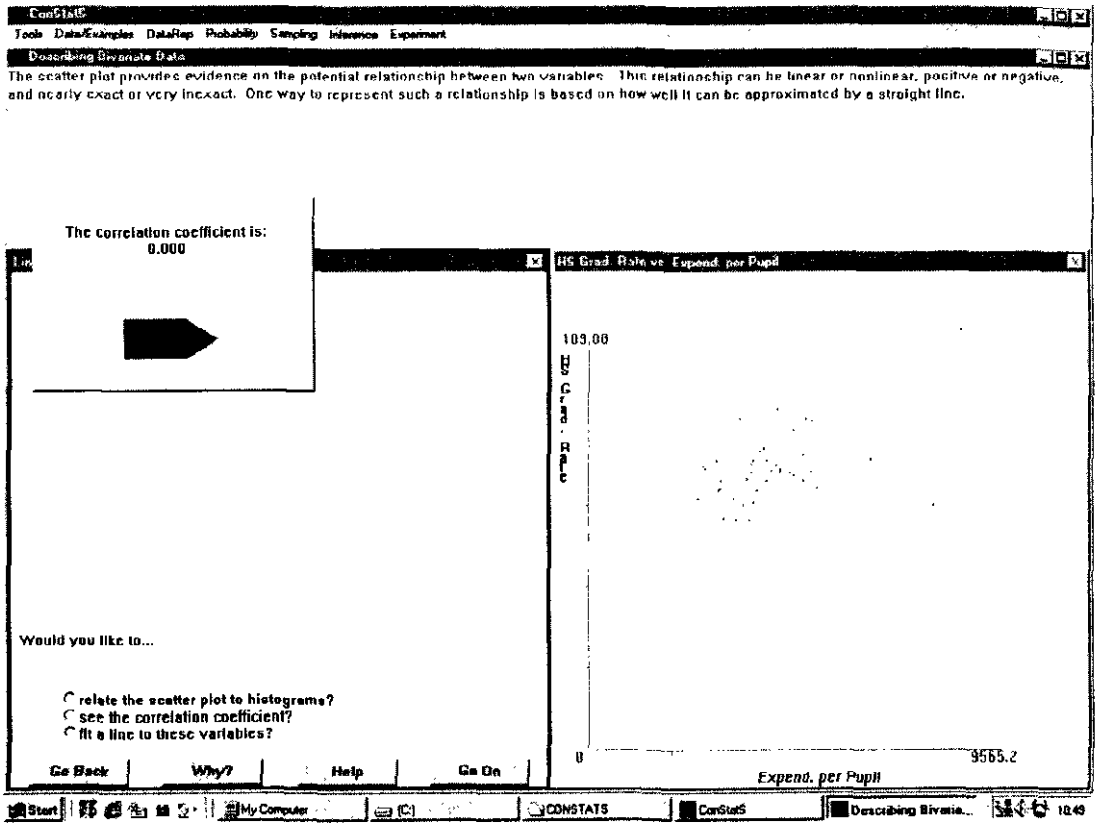
ConStatS has been described as Software for Conceptualizing Statistics (Tufts University Curricular Software Studio, 1997). This resource provides a program called *Representing Data* which is concerned with the different ways in which data are represented in statistics. A section of *Representing Data* is called *Describing Bivariate Data*, which introduces scatter plots and the correlation coefficient in the analysis of data. In this part of the program, a student can investigate how to display bivariate data on a scatter plot and also see the correlation coefficient for a data set. Having chosen a particular data set

that is available in *ConStatS*, such as the set called US Education, which provides data for high school graduation rates, teacher salaries and expenditure per pupil for each US state, a student can select 'Learn about bivariate techniques using this data set'. If a student selects this option, they can then select a pair of variables from the chosen data set and create a scatter plot, which is subsequently generated and displayed on the screen. For example, with the US Education data set, a user might choose to generate a scatter plot of high school graduation rate by expenditure per pupil for each US state. If a student chooses to move on in the program with these chosen variables, the following text is shown at the interface:

> "The scatter plot provides evidence on the potential relationship between two variables. This relationship can be linear or nonlinear, positive or negative, and nearly exact or very inexact. One way to represent such a relationship is based on how well it can be approximated by a straight line."

With respect to the displayed scatter plot of graduation rate and pupil expenditure a student can select the button 'see the correlation coefficient' and a pop up window is provided: "The correlation coefficient is: 0.000" (figure 4.2).

Figure 4.2 Describing Bivariate Data in ConStatS



In *Describing Bivariate Data* there is no mention of the *possible* values of the correlation coefficient, (or indeed, the different types of correlation coefficients that can be used to determine the direction and strength of relationship between variables), rather it is the detailed user manual (Smith, Cohen, Brown, Chechile, Garman, Cook, Ennis & Lewis, 1997) that describes the correlation coefficient as a numerical measure.

ConStatS was designed to be used by students from a variety of disciplines and the program provides generic and specific data sets relating to biology, sociology and economics. None of the available data sets relate specifically to psychology, but a user can create data sets, which can be used in the *ConStatS* environment. A data set can be used in *Describing Bivariate Data* to generate scatter plots and view the associated correlation coefficients. However, a student must rely on the *ConStatS'* user manual for essential details concerning the correlation coefficient as a measure of the direction and strength of the relationship. Although a student can select a variety of pairs of variables

from available data sets in *ConStatS* to generate and view scatter plots, they might not select pairs of variables in turn that could demonstrate the variety of relationships that can be obtained between variables. In the *ConStatS* environment no reference is made to the idea of correlation and causality. A student can carry out activities with data sets at the interface of *ConStatS*, (they can create a scatter plot and view the associated correlation coefficient), but they are not provided with specific feedback when they complete such tasks. For example, the student is not informed by means of a dialogue box that the scatter plot indicates that a very weak relationship exists between the selected variables.

4.5.2 *Introduction to Research Design and Statistics*

The demonstration version of *Introduction to Research Design and Statistics* (British Psychological Society, 1995) covers correlation. In the demonstration software that was reviewed, only two parts of the section *Correlation and Association* had been implemented. Specifically, the sub-section *What is correlation?* and the part *Pearson correlation* of the sub-section *Two types of correlation coefficient* were included in the software.

The sub-section *What is correlation?* provides a text-based outline of correlation and uses pictures, animation and every day examples, such as the relationship that tends to exist between waist measurement and weight to illustrate the idea of correlation (figure 4.3). Another screen describes how the degree of relationship between two variables can be expressed by the correlation coefficient (figure 4.4). Here, a learner activity is provided and a student can move the arrow on the slider that varies from -1 through 0 to 1. So, if a student changes the slider position to 0.8, the scatter plot that is also shown on the screen changes to represent this correlation.

Figure 4.3 What is correlation?

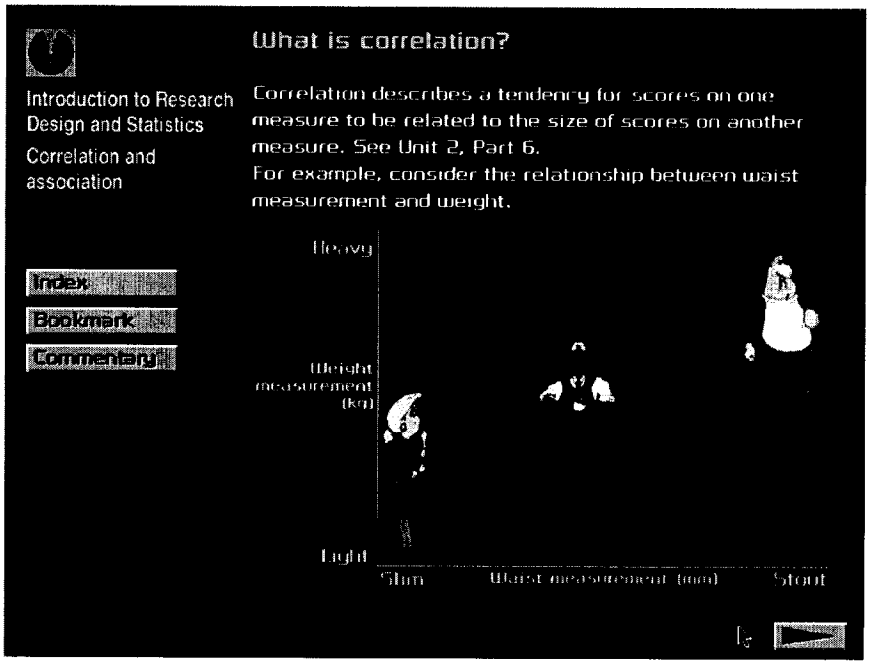
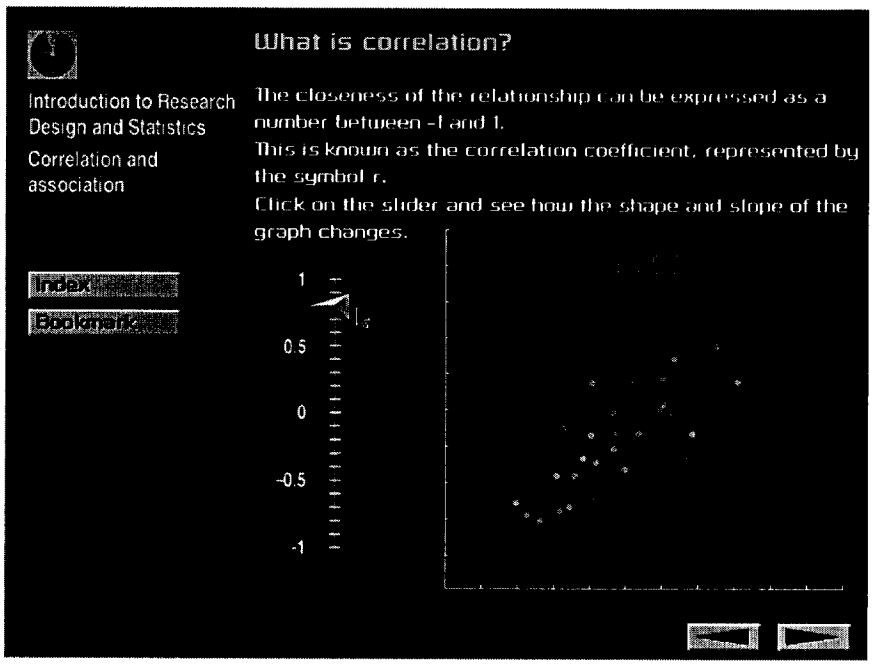


Figure 4.4 A learner activity in What is correlation?

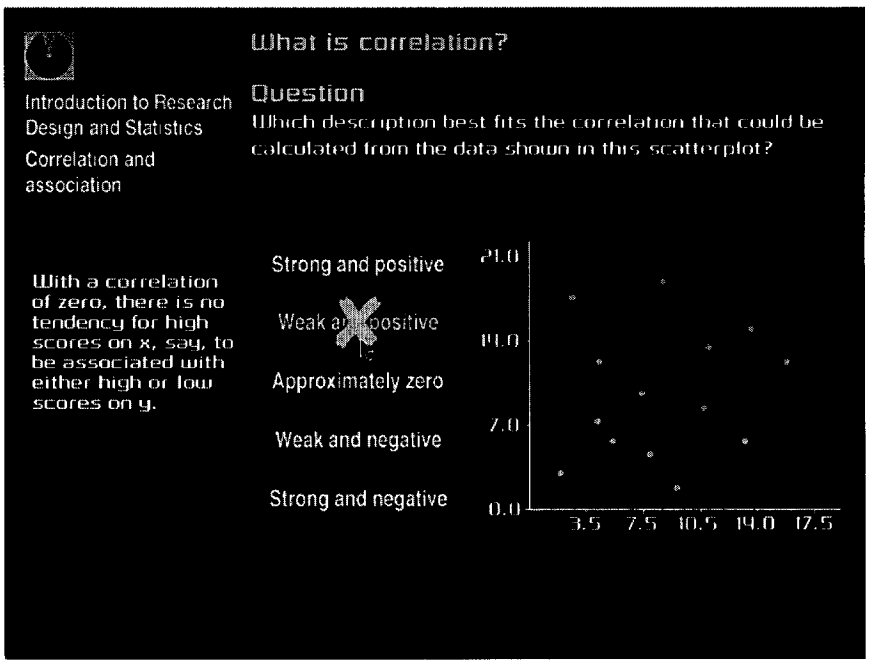


Having introduced correlation, *What is correlation?* offers a series of six questions for a student to answer. The feedback that is provided in response to a student answering a question is immediate, but limited because simple ticks and crosses are used to inform a user whether the answer(s) they have chosen are correct (figure 4.5). A 'Help' button on

the question screens can be selected by a student if they want a hint in answering the question correctly (figure 4.5). Three of the questions in the program are described here to illustrate what is covered by the demonstration disk.

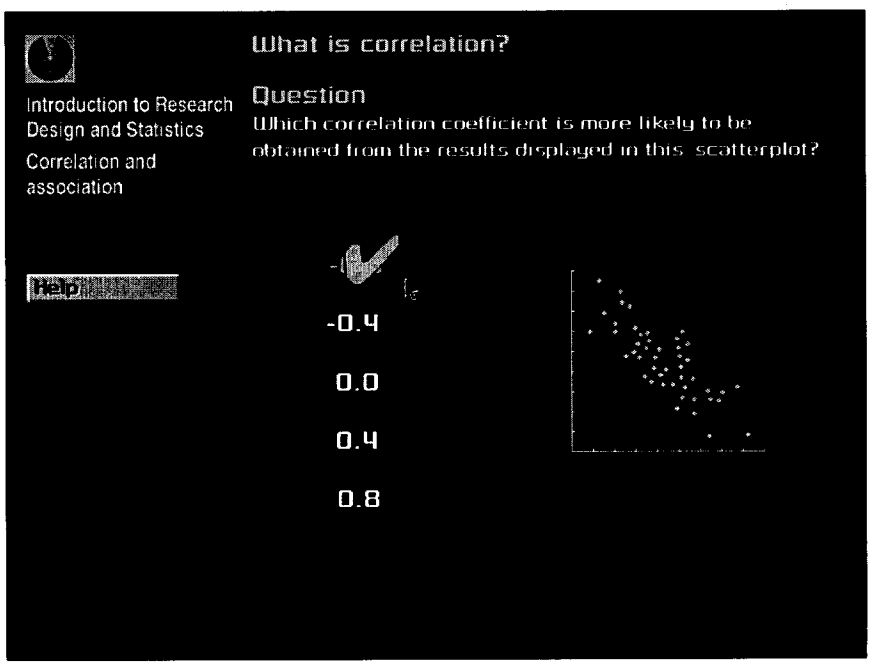
In one of the questions students are asked to match a description of a relationship with a displayed scatter plot (figure 4.5). In figure 4.5, the 'Help' button has been selected to provide a clue to answering the question correctly.

Figure 4.5 A student is asked to match the description of a relationship with a scatter plot



Another similar question asks which correlation out of a set of five coefficients is likely to be obtained from the results that are displayed in a scatter plot (figure 4.6). Here, a negative correlation is displayed by the scatter plot, and a student must select the correlation -0.8 to obtain a tick as feedback to the question.

Figure 4.6 Feedback to a question in What is correlation?

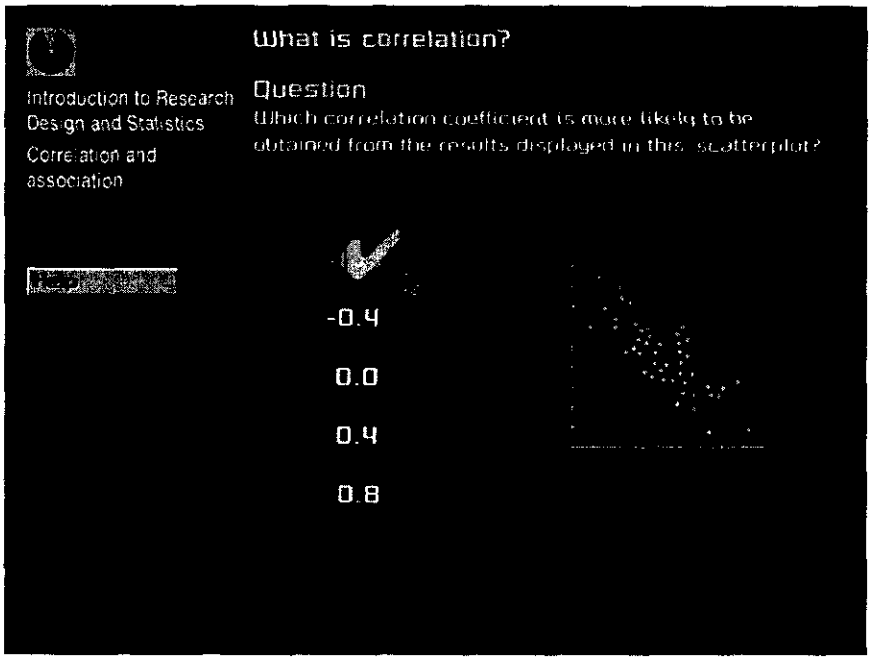


A student can also answer a question that asks which set out of four sets of scores represents the vertical axis of the scatter plot displayed on the screen.

In the part of the program called *Pearson correlation* a formula is animated so that it gradually appears on the screen. A student then has the option of clicking on a part of the formula for the Pearson correlation and accordingly, the screen runs through the calculation for this part of the formula. A student may, for instance, select $\Sigma x \Sigma y$ in the formula, and the screen will then provide a column of x and y data, show the sum of x and the sum of y for the two sets of scores and then multiply Σx and Σy. In this way, a student can be shown how to calculate parts of the formula to obtain the value of the coefficient. Alternatively, a student can simply select the arrow button that is displayed at the bottom of the screen and the computational procedure for calculating the coefficient will be demonstrated.

What is correlation? provides a useful summary for the concepts of positive, negative and zero correlation and the strength of correlations. A negative correlation is clearly defined by using an every day example. A learner activity and one of the questions

Figure 4.6 Feedback to a question in What is correlation?



A student can also answer a question that asks which set out of four sets of scores represents the vertical axis of the scatter plot displayed on the screen.

In the part of the program called *Pearson correlation* a formula is animated so that it gradually appears on the screen. A student then has the option of clicking on a part of the formula for the Pearson correlation and accordingly, the screen runs through the calculation for this part of the formula. A student may, for instance, select $\sum x \sum y$ in the formula, and the screen will then provide a column of x and y data, show the sum of x and the sum of y for the two sets of scores and then multiply $\sum x$ and $\sum y$. In this way, a student can be shown how to calculate parts of the formula to obtain the value of the coefficient. Alternatively, a student can simply select the arrow button that is displayed at the bottom of the screen and the computational procedure for calculating the coefficient will be demonstrated.

What is correlation? provides a useful summary for the concepts of positive, negative and zero correlation and the strength of correlations. A negative correlation is clearly defined by using an every day example. A learner activity and one of the questions

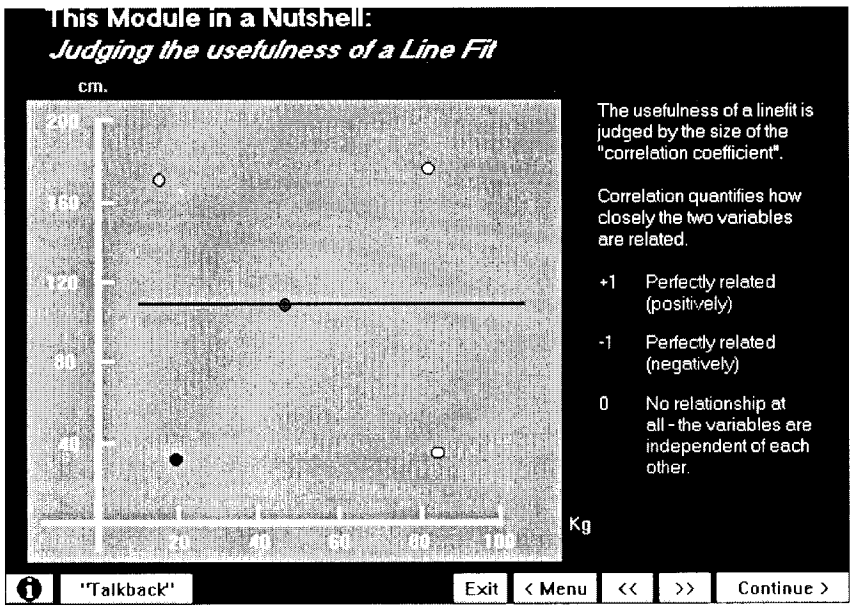
in the program show how the value of a correlation coefficient and a scatter plot can describe a relationship between two variables: in the learner activity a student can change the value of the correlation coefficient and view the change of pattern on the scatter plot and in one of the questions they can choose which coefficient matches the pattern on a displayed scatter plot. However, causality and correlation are not explicitly covered in the program, and although the set of six questions are varied in content and format, only limited feedback to these questions is provided in the form of ticks and crosses for correct and incorrect answers.

Given that data-analysis software can now be used by psychology students to calculate statistics, such as correlation coefficients, it does not seem appropriate that a procedure derived from a formula to calculate a statistic is demonstrated in a computer-assisted learning program.

4.5.3 *Statistics for the Terrified*

The computer-assisted learning program, *Statistics for the Terrified* (Morris, Szuscikiewicz & Preston, 1995) includes the module *Fitting Lines to Data* that covers correlation. Rather unconventionally, in this program, correlation is introduced in the context of regression analysis. On one of the screens, the student is simply informed by means of a scatter plot and text that a correlation measures how closely two variables are related so that when the text "+1 perfectly related (positively)" is first displayed on the screen, the scatter plot displays the data points as a perfect line, and subsequently the pattern of the scatter plot changes to illustrate a perfect negative relationship and no relationship (figure 4.7).

Figure 4.7 The correlation coefficient is introduced in Statistics for the Terrified

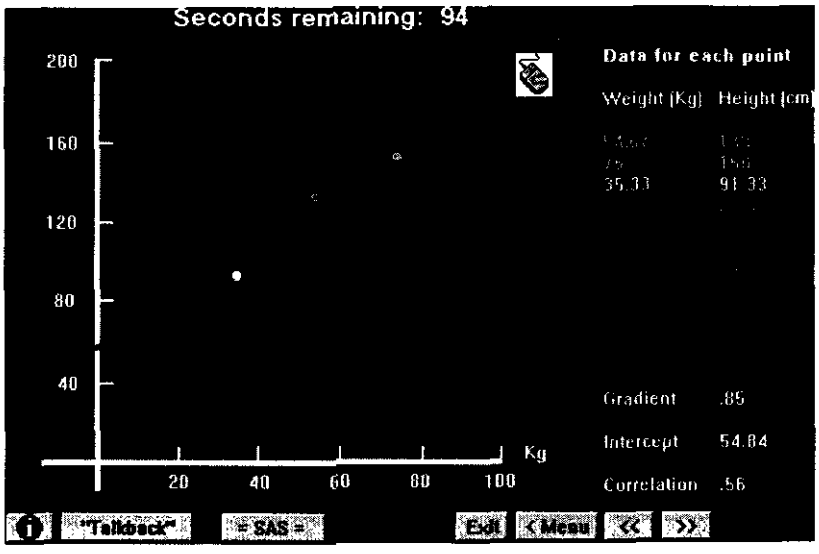


Much of the module is concerned with the details of fitting lines to bivariate data, but there are three more screens that pertain to the concepts of correlation. One of these screens provides an exercise for a student to carry out in a specified period of time (figure 4.8). A student is provided with a data set of weight and height that is represented by a scatter plot on the screen and the value of the correlation coefficient. Prior to this exercise, the student is informed that

> "For the next two minutes you can drag any of the ten points to new locations on the graph."

Indeed, by using the mouse a student can select a data point on the scatter plot and move it, and the individual values of the data set for weight and height and the value of the correlation coefficient change accordingly. It is not entirely clear why a period of two minutes is given for students to play around with these data points.

Figure 4.8 An exercise with 94 seconds remaining



Fitting Lines to Data also provides a screen of text which simply summarises correlation (figure 4.9), but a student can complete an exercise in which they move the data points of a scatter plot to obtain a correlation of, for example, 0.8 (figure 4.10). While a student selects and drags certain data points they will see the correlation coefficient change and they can directly manipulate the points until the required correlation is obtained. If a student attempted this exercise on another occasion, the module is set to ask a user to try and obtain a different value of a coefficient by dragging the data points on the scatter plot.

Figure 4.9 Correlation is summarised in Statistics for the Terrified

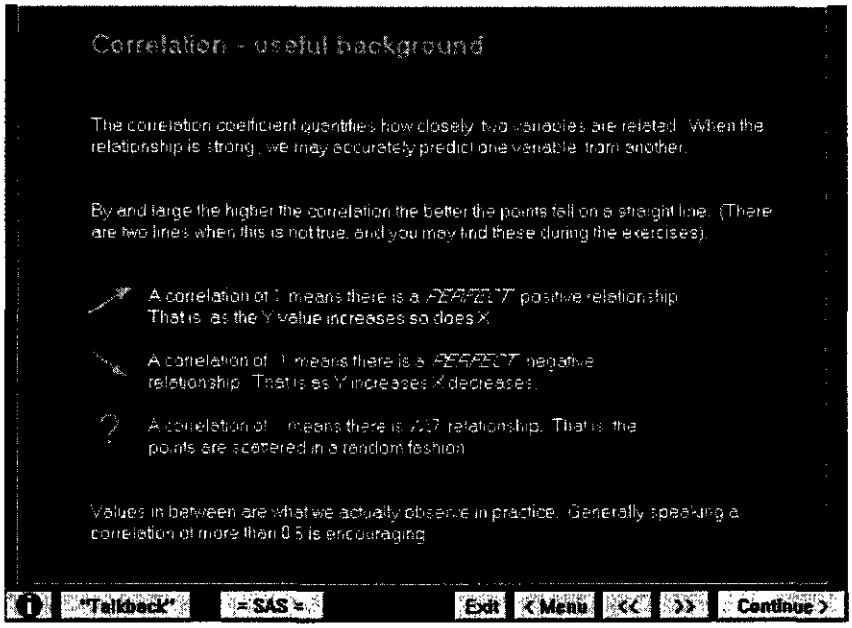
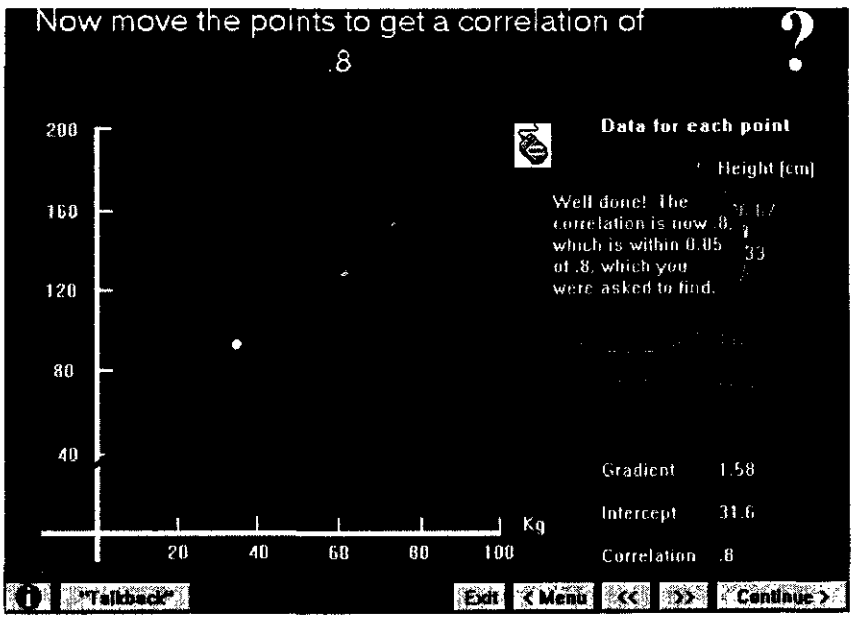


Figure 4.10 An exercise that involves moving data points on a scatter plot



In the parts of the module that concern correlation, only one set of bivariate data (weight and height) is used to illustrate possible relationships. It is also a concern that when correlation is summarised on one of the screens (figure 4.9) it is stated that

"The correlation coefficient quantifies how closely two variables are related. *When the relationship is strong, we may accurately predict one variable from another*" (emphasis added).

This kind of wording could confuse a learner because it might suggest to them that one variable is causally related to another variable. In addition, it might not be appropriate to firstly obtain a correlation from data and then to go about predicting one variable from another variable because many empirical studies in the social sciences use a correlational design where variables are not experimentally controlled. The *Fitting Lines to Data* module does not mention correlation and causation or the care that must be exercised in the interpretation of a correlation that indicates a strong relationship.

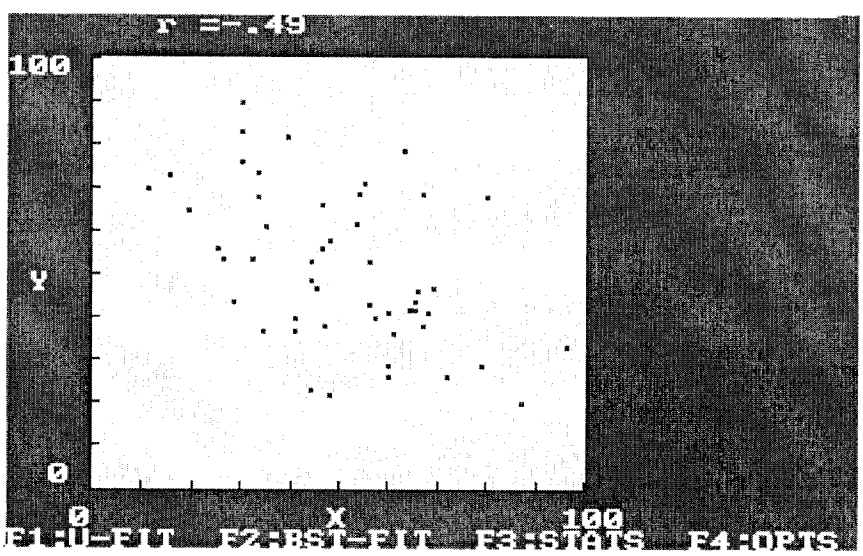
As described above, *Statistics for the Terrified* does offer two interesting activities concerning correlation that a student can complete. In these activities a student is given useful feedback either in terms of the changing values of a data set and the correlation coefficient, or in terms of performance related feedback with respect to whether a student manages to obtain a specified value of a correlation by dragging points on a scatter plot. However, the activities for correlation are limited in terms of frequency and variety.

4.5.4 *Statistics Tutor: Tutorial and Computational Software for the Behavioral Sciences*

Statistics Tutor was designed "to complement any course in statistics and design offered at the undergraduate level" (Allen & Pittenger, 1991, p. vii). *Statistics Tutor* includes a part called *Correlation Tutorial*, but uses MS-DOS and so the human-computer interface of the program has a rather out of date look (figure 4.11). The user manual for *Statistics Tutor* has a section corresponding to the *Correlation Tutorial*, and this manual provides descriptions of the relevant statistical concepts and exercises to carry out with the tutorial. However, the *Correlation Tutorial* is very limited and only offers one exercise that can be carried out in three different ways. The main menu screen of the tutorial simply offers

three options for a student: 'scatter plot via computer', 'scatter plot via keyboard' and 'scatter plot via data file', providing three methods for producing scatter plots. If a student chooses to generate a scatter plot via the computer, they are provided with an empty scatter plot and are prompted to enter a value for a population correlation coefficient called *rho*, (which is symbolised as ρ), and the number of data pairs to be plotted. So, with this exercise the student can experiment with different sample sizes and different values of population coefficients. If a student entered ρ as -0.6 and a sample size of 50, the Pearson correlation coefficient, r, is calculated for a sample generated from the specified population and displayed above the corresponding scatter plot (figure 4.11). Here, the correlation coefficient calculated from the sample that was generated by the program is -0.49. It is pointed out in the user manual that the computer simulates sample sets of correlated scores providing estimates of the population correlation coefficient. With this exercise, a student may come to recognise that the characteristics of a sample more nearly reflect the characteristics of the population from which it is drawn as sample size increases.

Figure 4.11 An exercise from the Correlation Tutorial



As described above, there is only one kind of exercise provided for a student in the *Correlation Tutorial*. This activity might assist students' understanding of the

importance of sample size in sampling from a given population. With the activity students also view the value of a correlation coefficient and the corresponding scatter plot. The activity, however, is limited because it does not directly impart concepts concerning correlation by, for example, stating that a correlation is an index of strength for a relationship between variables and relating this to the correlation coefficient obtained for a sample. The idea of correlation and causation is not addressed by *Statistics Tutor.*

4.5.5 *STEPS*

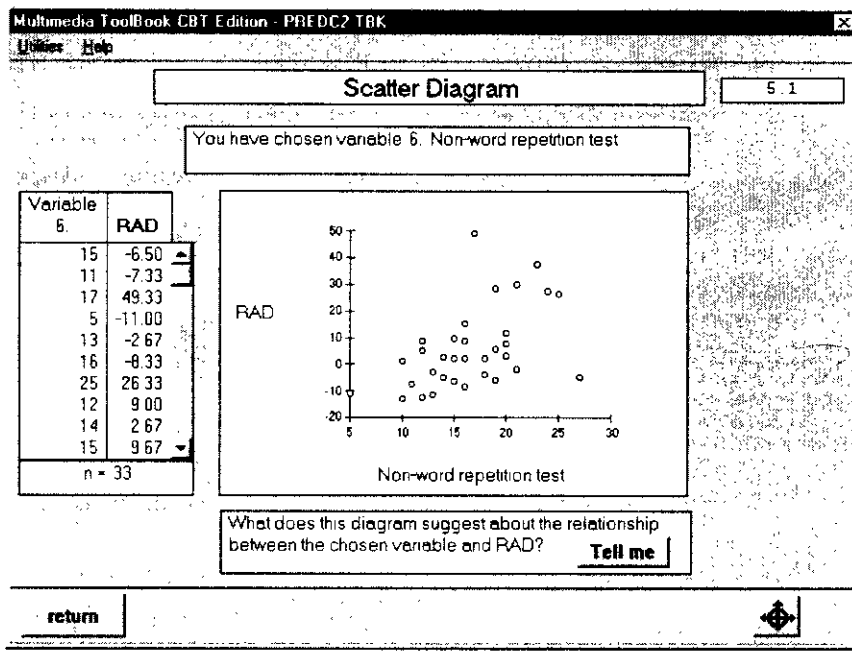
As part of the *STEPS* project (Statistical Education through Problem Solving, 1996), computer-based learning materials, which have been designed to teach the use of statistics in psychology, have been developed. The computer-assisting learning program *Predicting Dyslexia?* covers correlation and presents educational objectives on one of the introductory screens as follows

> "On completing this module you will:
>
> • be able to interpret scatter diagrams
>
> • know how the correlation coefficient measures strength of association between variables
>
> • have calculated a correlation coefficient."

Predicting Dyslexia? introduces correlations in the context of data that was obtained from a sample of pre-school children. This data set is used in the program to illustrate scatter plots and correlations and a student can view the whole of the data set, which consists of children's scores on a variety of tests. The data set includes children's scores on a vocabulary test, a Non-word Repetition test, rhyming tests and a Reading Age Deficiency score (RAD). It is not stated in the program or in the accompanying *STEPS* documentation whether the data set used in the program is genuine and from a real study in psychology. The program provides details of the study and descriptions of the

variables in the form of text exposition. On a screen titled 'Describing relationships using correlation', a student can select one of the sets of scores to investigate its relationship with RAD. For example, the student can select the variable Non-word Repetition test (and RAD is always the other variable), and they can then either select the button 'Scatter diagram' or the button 'Correlation'. The former invokes a screen that shows the values for the pair of selected variables and the corresponding scatter plot (figure 4.12).

Figure 4.12 A scatter plot can be generated in Predicting Dyslexia?

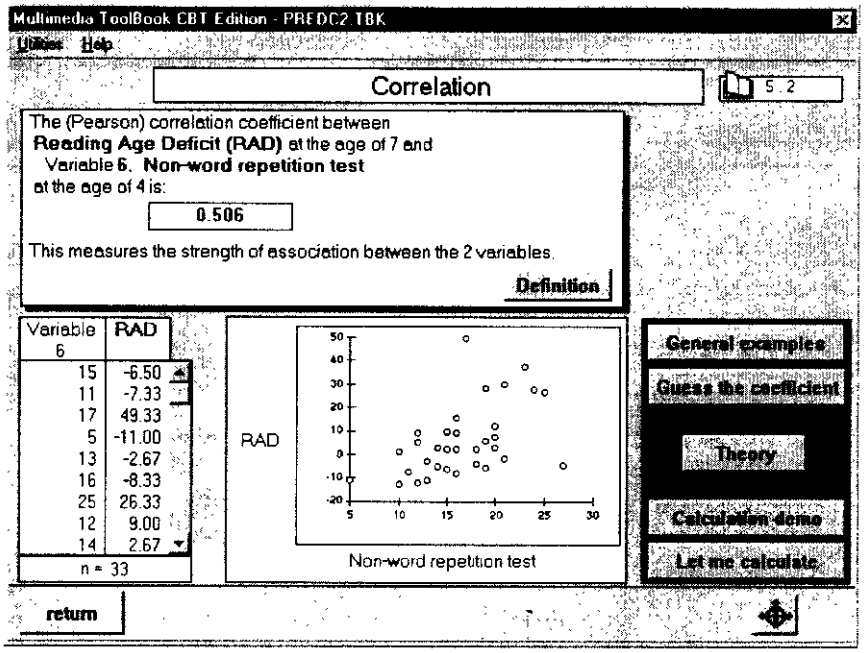


A student can therefore select a number of different variables from the data set to see how they correlate with RAD and view the correlation on a scatter plot. It is a pity that all the possible pairs of variables give a relationship that is positive. In fact, all of the correlations between RAD and the tests in the study are between 0.5 and 0.6. So, if a user selects the button 'Tell me', they always get feedback that states "as the variable's value increases, the value of RAD also tends to increase."

Having chosen which variable to pair with RAD (e.g., Non-word Repetition test), a student can select the button 'Correlation' which invokes a screen showing the correlation coefficient obtained between the two variables in question. This screen also

shows the corresponding scatter plot and the values for the chosen variables (figure 4.13). From here, a student can move on to look at 'General examples of correlation' or to 'Guess the correlation coefficient'.

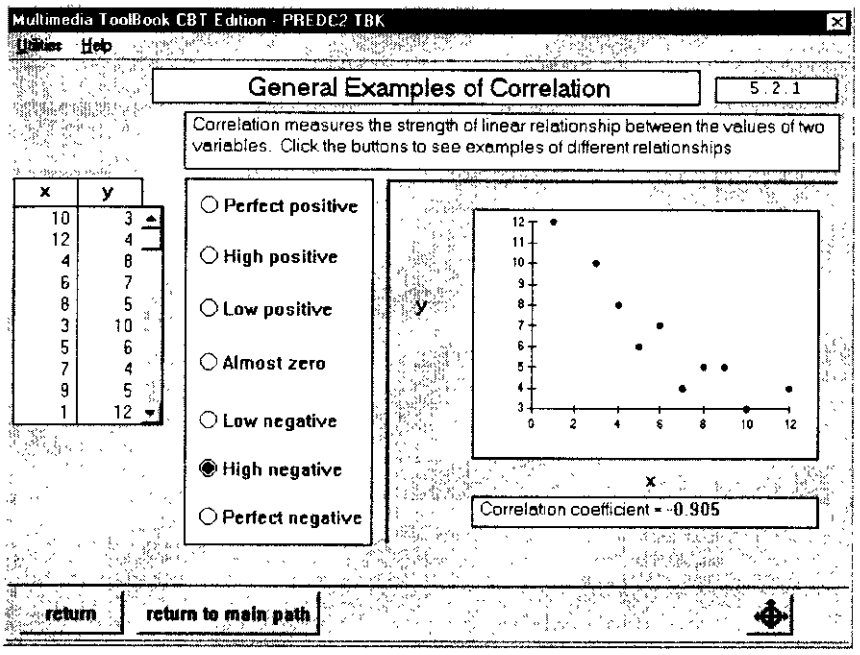
Figure 4.13 A correlation coefficient is displayed for a pair of variables



On the screen 'General examples of correlation', a student is presented with a list of radio buttons to choose from which are titled: Perfect positive, High positive, Low positive, Almost zero, Low negative, High negative and Perfect negative (figure 4.14). A table with x, y columns, a scatter plot, and a box called the correlation coefficient that displays a value (e.g., - 1.000), is also presented on the screen. All these representations are linked and change dynamically. For example, if a student selects the radio button titled 'High negative' the graph changes displaying a pattern for a negative correlation, the value -0.905 is displayed, and the data in the table also change (figure 4.14). However, a student can only change the scatter plot and associated data values by selecting the radio button options and on doing this they are not informed what the changing values mean in terms of the data. For example, feedback is not given that specifies that as one variable generally increases, then the other variable decreases. It is surprising to find that this

activity does not use the data set concerning children's scores on tests that is used elsewhere in the program.

Figure 4.14 General examples of correlation



On the screen 'Guess the correlation coefficient', students can estimate the correlation coefficient that is displayed on a scatter plot. Here, a student can also change the data by selecting one of three available data sets or by dragging the data points on the scatter plot to different positions. If a student changes the pattern of the scatter plot or selects one of the three available data sets, the data values in the x, y table change accordingly. On this screen, a student can have a go at entering the value of the correlation coefficient that corresponds to the data displayed on the scatter plot. Once a student has guessed the correlation, they are provided with appropriate feedback that either informs them that the value is too low or too high or that they have guessed correctly. Alternatively, a student can select 'Display actual coefficient' and the correct value of the coefficient is displayed on the screen.

Predicting Dyslexia? includes a screen that provides a demonstration of the calculation of a correlation coefficient, but this does not use the data set concerning

children's reading and cognitive abilities, which are meant to provide a context for the application of statistical concepts and techniques introduced in the program. Another screen in the program has been designed to allow students to calculate a correlation coefficient and here also, bivariate data from the program's data set are not used. Rather, the 'Let me calculate' exercise uses small data sets for convenience and primarily consists of having students insert two missing items in the data table set up for computation of the correlation coefficient. A student can also input worked out values in the formula that is provided, or they can simply select the button 'Do it for me'. In the case of the latter, the program runs through the calculation of a correlation by inserting the appropriate sub-calculations and values in the formula to give the answer.

The *STEPS* computer-based learning materials for psychology are meant to be based around specific problems arising in this discipline. So, the *Predicting Dyslexia?* module uses a study concerning dyslexia and the associated data set. Correlation and regression techniques are used in the program to see how well the different tests predict the children's RAD scores. This is rather nice. However, it is not stated in the program whether the data set is from genuine psychological research, but more importantly, four of the activities that concern correlation do not use this data set. As described above, there are two activities in the program that do use the data set: a student can select bivariate data from the data set and subsequently view the corresponding coefficient, or the relationship on a scatter plot. Yet this activity is limited by the fact that all possible pairs of variables, which may be selected by a student, have a positive correlation of between 0.5 and 0.6. Bivariate data is required so that students can view negative correlations and very weak correlations on a scatter plot. *Predicting Dyslexia?* makes no mention of causation and correlation and none of the activities address this statistical issue.

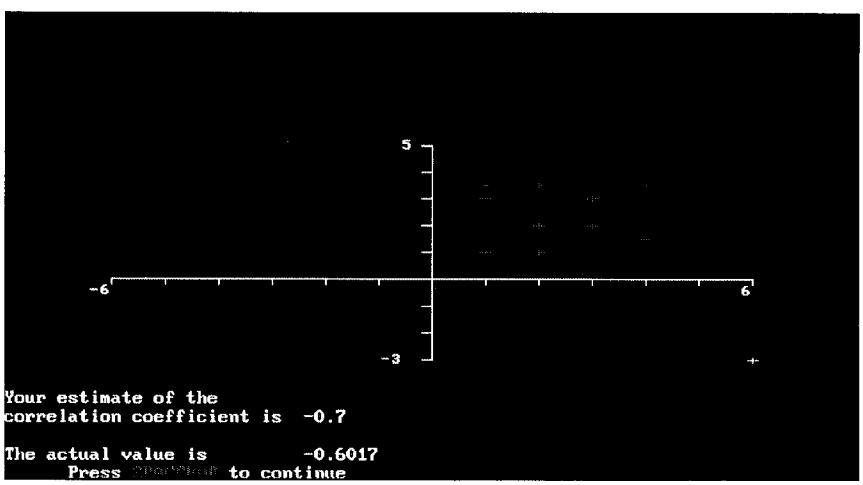
4.5.6 *Understanding Statistics*

The software *Understanding Statistics* (Royal Statistical Society, 1996) includes a program covering correlation. This program offers two exercises relating to correlation: 'A first look at correlation' and 'Display data graphically'. In the former, a student can estimate the correlation coefficients of scatter plots. The aim of this exercise it to give a learner

> "a feel for the product moment correlation coefficient for different types of sample and ... to illustrate how an extra outlying point can radically affect the value" (Barnett & Holmes, 1996, p. 55).

On this exercise, the data sets are displayed as scatter plots and a student can enter an estimate of the value of the correlation coefficient. The program displays a sequence of eight scatter plots and a student can estimate the value of the correlation for each in turn. For example, if a student types the estimate of -0.7 for one of the scatter plots that contains an outlier, they are provided with the following feedback: "The actual value is -0.6017" (figure 4.15). This is rather uninformative. If a student simply selects 'Return' on this exercise, then the true value of the coefficient is given before an estimate is made by a student.

Figure 4.15 An exercise from Understanding Statistics



A student can load a data file and choose to 'Display data graphically'. This option simply displays the data as a scatter plot and gives the values of three different types of correlation coefficients, which can be slightly different for a given set of data. Neither the manual or the program explain why different correlation coefficients are used for bivariate data.

Understanding Statistics only provides two activities for correlation. For one, a student can estimate the correlations for a series of scatter plots. Although this activity could illustrate to a learner the effect outliers have on the correlation coefficient (figure 4.15), the feedback to this activity is limited and simply gives a student the value of the correlation coefficient to four decimal places. In addition, the series of eight scatter plots is fixed so that each time a student uses the software they are presented with the same set of scatter plots. The second activity simply involves the generation of a scatter plot and the calculation of three different correlation coefficients for a set of data. This is a data-analysis activity that makes little sense because a particular correlation coefficient is usually chosen depending on how variables have been measured. The statistical issue of correlation and causation is not addressed by *Understanding Statistics*.

4.6 Discussion

At the beginning of the chapter, it was highlighted that constructivist conceptions of learning have described learning as cumulative and that the acquisition of concepts is facilitated if they are anchored in realistic contexts or problems. It was therefore proposed that research studies in psychology should be used to provide a context to present and illustrate statistical concepts to psychology students. *Stat Lady* was designed to make the learning of statistics more meaningful and therefore more memorable by introducing concepts and principles in the context of realistic problems. Evaluation studies of *Stat Lady's* probability module have indicated that significant learning gains were made by students who used this module, but equivalent gains were also made by students who attended a series of lectures or by students who completed a Workbook version of the

Stat Lady curriculum. A major concern was raised concerning the approach taken in the design of *Stat Lady* because research related to students' understanding of probability has not been addressed by Shute et al (1996). *Stat Lady* was not designed to address specific probability concepts that students find difficult to understand.

It is widely recognised that a learner's prior knowledge can affect the acquisition of new knowledge: it might impede or facilitate the acquisition of new material (O'Shea, 1992; Shuell, 1992). A learner's prior knowledge in the form of, for example, prior conceptions or misconceptions must therefore be addressed by instructional materials. *StatPlay* was designed to help students overcome statistical misconceptions by using demonstrations and dynamically linked multiple representations (Cumming & Thomason, 1998). Further empirical work is required to investigate whether *StatPlay* contributes to students' understanding of those statistical concepts that the program was designed to address.

A review of computer-assisted learning programs that cover correlation was undertaken to see what kinds of learner activities could be used in a program and to identify limitations of existing programs. This review of six programs raises six issues, which will be discussed in turn, relating to the design of a computer-assisted learning program for correlation.

Firstly, the reviewed programs used data sets, familiar examples or variables x and y to present correlations. Only two of the programs (*ConStatS* and *Predicting Dyslexia?*) use data sets, from which a user can select bivariate data, to introduce and illustrate the concepts of correlation. In *ConStatS*, bivariate data from say, the US Education data set can be selected to generate a scatter plot and a correlation coefficient. Although data sets can be created to be used in *ConStatS*, none of the program's existing data sets relate directly to psychology. *Predicting Dyslexia?* was designed to present statistical concepts and techniques in the context of a problem in psychology. The problem is to see what variables, in the form of a variety of test scores, can predict RAD.

Predicting Dyslexia? describes the variables in a research study in psychology and a user can choose to examine the whole data set or select different pairs of scores to see what relationships exist between different variables. However, it is not clear whether the research study and the corresponding data set are genuine. In addition, the learner activities in the program do not always use the data set that is the focus of *Predicting Dyslexia?*

Introduction to Research Design and Statistics and *Statistics for the Terrified* both use familiar examples, such as weight and height to illustrate correlations. In *Statistics Tutor* and *Understanding Statistics* variables are simply known as x and y.

Secondly, all of the computer-assisted learning programs provide learner activities where a student can carry out an action (or a series of actions) at a program's human-computer interface. Every one of the programs provides learner activities that link a correlation coefficient with a corresponding scatter plot. In *ConStatS*, a student can select bivariate data and generate the scatter plot and correlation coefficient for this data and can view both simultaneously. Similarly, in *Predicting Dyslexia?* a student can select a set of scores to pair with the variable RAD and subsequently view the corresponding scatter plot and coefficient. In *Introduction to Research Design and Statistics*, a student can select and drag an arrow on a sliding scale, which varies from -1 to 1, to see that changing the value of a coefficient will change the pattern of data on a scatter plot. A learner can move the data points on a scatter plot that is provided in *Statistics for the Terrified*. Here, they must try to select and position the data points to obtain a specified correlation coefficient.

In *Statistics Tutor* only one learner activity is provided where a student enters a value for the population coefficient and a sample correlation coefficient is generated from the population and displayed on the screen with the associated scatter plot. A student must estimate the value of the correlation for a series of scatter plots in *Understanding Statistics*. By completing this activity, students are told the exact value of a coefficient for a given scatter plot and therefore view both concurrently.

Thirdly, the learner activities in the programs provide feedback and this must be considered. Four of the six programs offer feedback to a student's actions at the interface that is limited (*ConStatS*, *Introduction to Research Design and Statistics*, *Statistics Tutor* and *Understanding Statistics*). For example, *Introduction to Research Design and Statistics* provides a variety of questions for the student to answer, but only ticks and crosses are used to indicate correct and incorrect answers. The feedback to a learner activity in the *STEPS* module, *Predicting Dyslexia?* is a little more informative. When a student has guessed the value of a correlation coefficient from a scatter plot, they are provided with appropriate feedback that states whether the entered value is too high or too low.

Fourthly, the examined computer-assisted learning programs do not necessarily address the misconceptions concerning correlation that students hold. *ConStatS*, *Introduction to Research Design and Statistics*, *Statistics for the Terrified*, and *STEPS* do cover the concepts of negative correlation, positive correlation and no correlation, and use text, correlations coefficients and scatter plots to present these concepts. *Statistics Tutor* and *Understanding Statistics* do not cover these concepts explicitly. The strength of correlations is, however, only considered in two of the programs: *Introduction to Research Design and Statistics* and *Predicting Dyslexia?* The statistical issue of correlation and causation, and the possible interpretations of an obtained correlation is not covered by any of the computer-assisted learning programs. *Introduction to Research Design and Statistics* is, however, a demonstration version, which has a section called Correlation and Causality that had not been implemented.

Fifthly, two of the programs look at statistics formulas that can be used to calculate a correlation coefficient. *Introduction to Research Design and Statistics* provides a demonstration of how to calculate parts or the whole of a formula to obtain a coefficient. *Predicting Dyslexia?* also provides a demonstration of how to calculate a coefficient and an activity that gives a student the opportunity to practice calculating parts of a formula. In

the study reported in chapter 3 it was found that students had difficulties in using a formula to obtain a statistic. It was considered that this finding indicated that students lacked the prerequisite mathematical skills to derive a procedure from a statistical formula. However, it can be argued that data-analysis software can now be used by psychology students in higher education to generate scatter plots and correlation coefficients for data sets. It is therefore recommended that a computer-assisted learning program should not present computational procedures to be learnt by students so that they can calculate statistics.

Finally, there are advanced topics relating to the topic of correlation that are addressed by *Statistics Tutor* and *Understanding Statistics*. In *Statistics Tutor*, a student can specify the population coefficient and a sample size and the program generates bivariate data from the given population. The sampled data is shown on a scatter plot along with the corresponding correlation coefficient. From this activity, a student might come to appreciate the important role of sample size in sampling from a given population. There are three scatter plots in *Understanding Statistics* that have outliers. An outlier on a scatter plot will influence the size of a correlation coefficient as described in chapter 2. The effect of outliers on a correlation is, however, a more complicated methodological issue than basic relationships between variables (Goldstein & Strube, 1995). The influence that outlying points have on the value of a correlation coefficient could be looked at by students after the main concepts of correlation have been understood.

4.7 Summary

The previous chapter showed that psychology students do hold statistical misconceptions that concern correlation and it is these that must be addressed in the design of a computer-assisted learning program. This chapter considered theoretical perspectives that have proposed that learning is cumulative and is facilitated if concepts are anchored in realistic contexts or problems. These general learning principles informed the design of *Link*, which is described in the following chapter. Research and developments in the field of

computer-assisted learning for statistics were reviewed in this chapter. In addition, a review of a variety of computer-assisted learning programs that cover the topic of correlation was described. The design and implementation of *Link* is described in the following chapter. Both formative and summative evaluation studies of *Link* were conducted and the methodology employed for these studies is considered in chapter 6.

Chapter 5

The design and development of *Link*

5.1 Introduction

This chapter describes the design of *Link*, which was informed by research-based principles of learning, empirical work concerning students' statistical misconceptions, research and developments in the field of computer-assisted learning for statistics and a review of programs that cover correlation. The design of *Link* was also informed by recommendations that have been made that relate to the effective design of computer-assisted learning programs (e.g., Shuell, 1992).

The approach taken in the development of *Link* is also described in this chapter. The authoring tool Macromedia Director 5.0 was used to develop a first prototype of *Link*. This is illustrated here by user scenarios of how a student would work through the program and complete the different learner activities provided. The methodology that was employed in the evaluation studies of *Link* is described in the following chapter. As part of program development, a formative evaluation and an expert evaluation of the program were carried out and these will described in chapter 7 and chapter 8.

5.2 The design of *Link*

If effective computer-assisted learning programs are to be designed then the target population of the application must be considered. Shuell (1992) has emphasised that in the development of instructional computing systems the audience of the program must be considered, which means that student characteristics should be taken into account when developing a system. Furthermore, the development cycle of a computer-assisted learning program must involve early testing of the application with target users. It is imperative that in the development of a system the prior knowledge of the target population of students who would use the program is considered (Shuell, 1992). Shuell (1992) has

pointed out that:

> "developers also need to find ways of diagnosing relevant student characteristics and determining the range of knowledge that the students might exhibit" (ibid., p. 46).

Chapter 3 showed that psychology students have particular difficulties and confusions relating to correlations and it is important that the design of a computer-assisted learning program for this area must address such problems. More specifically, the findings of the investigation that were described in chapter 3 indicated that students have confusions relating to negative correlations, the strength of correlations and causality. The findings were discussed in relation to empirical work that has looked at students' misconceptions concerning correlation (Batanero et al, 1997). Three misconceptions in particular have been identified.

- Causalistic conception.

This conception describes the idea that given a correlation, a student thinks that one variable, A has a direct causal influence on another variable, B and he/she does not entertain any other possible interpretations of the correlation (e.g., B could also have a causal influence on A or a third variable could be responsible for the correlation).

- Unidirectional conception.

This conception describes one or more or the following: (a) a negative correlation coefficient is thought to indicate no correlation, (b) a negative correlation that is displayed on a scatter plot is viewed as a relationship or as a positive relationship, (c) a negative correlation that is displayed on a scatter plot is viewed as indicating no relationship between variables.

- The conception that a positive correlation is stronger than a negative correlation.

A student thinks that a positive correlation coefficient is stronger than a negative correlation when this is not the case.

Drawing on theoretical perspectives (e.g., Bransford et al, 1990), in chapter 4 it was proposed that the acquisition of statistical concepts is facilitated if they are presented to psychology students in the context of a psychological study. *Stat Lady* was designed to facilitate the learning of statistics by situating or anchoring concepts and principles in realistic scenarios (Shute et al, 1996). In a similar way, the *STEPS* module *Predicting Dyslexia?* was designed to illustrate statistical concepts and techniques in the context of a psychological study that presented a research problem. In the module, a user can investigate the relationships that exist between different variables in the study. A research study in psychology could provide a context for the learning activities in a computer-assisted learning program.

In the light of the above, it was decided that a study from psychology concerning TV violence and children's aggression would be used in the first prototype of *Link*. A study concerning TV violence and children's aggression (Eron et al, 1972) was used for question 17 in the investigation that was described in chapter 3. An adapted version of this study provided a context to present the learner activities in the program.

It was emphasised in chapter 4 that learning is cumulative and that the effective design of instructional materials should address students' prior knowledge in the form of, for example, misconceptions. *StatPlay* was designed to address students' misconceptions in statistics that are resistant to traditional forms of education (Cumming & Thomason, 1995) Cumming and his associates (1995, 1998) have argued that *StatPlay* helps students to overcome their statistical misconceptions because it provides demonstrations and dynamically linked multiple representations of statistical concepts. There has, however, been very little work on whether particular kinds of learner activities in a computer-

assisted learning program address students' misconceptions. For example, further evaluations of *StatPlay* are required to see whether dynamically linked representations of statistical concepts do remedy students' misconceptions. One of the main objectives of developing *Link* was to investigate whether particular computer-based learner activities address students' misconceptions concerning correlation.

In chapter 2, it was emphasised that although much research has looked at students' misconceptions, there has been a lack of research that has investigated whether particular learning conditions affect misconceptions. However, Driver (1988) has described a number of teaching strategies that were used to facilitate the construction of new concepts, which have been used in teaching sequences for topic areas in science. The strategies that are used in *Link's* learner activities are similar to three of these teaching strategies: 'broadening the range of application of a conception', 'differentiation of a conception' and 'the construction of an alternative conception' (Driver, 1988). The design of these learner activities is described in the following section (see 5.3 *Link*).

A review of computer-assisted learning programs that cover correlation indicated that a program for correlation should:

- Link correlation coefficients with scatter plots.
- Provide learner activities with informative feedback.

In general, the reviewed programs covered the concepts of negative correlation and no correlation by providing learner activities that involved the presentation of scatter plots and correlations coefficients simultaneously, and which therefore demonstrated a variety of relationships to a learner. However, only two of the programs addressed the strength of correlations and none of the examined programs addressed the statistical issue of correlation and causation.

In the light of the review, it was argued that a computer-assisted learning program should not provide a formula to calculate correlations, a demonstration of how to calculate

correlations from a formula, or an opportunity to practice the calculation of a correlation coefficient. This is because the statistics curriculum is changing with the increasing availability of data-analysis software and so psychology students do not necessarily have to learn how to calculate a statistic from a computational procedure derived from a formula.

There are programs that demonstrate the effect that outliers have on the value of a correlation coefficient (e.g., *Understanding Statistics*), but this methodological issue was not addressed by *Link* because students' understanding of the influence of outliers on a correlation was not investigated in the empirical study that was outlined in chapter 3. This topic is also considered to be a more complicated topic for students than that of basic relations between variables (Goldstein & Strube, 1995).

Link was therefore designed to include learner activities, which use data in the form of correlation coefficients and scatter plots, and which provide informative feedback to a student. These learner activities were designed to address the misconceptions concerning correlation that were outlined above. Specifically, as will be described below, the first prototype provides a learner activity for each of the three misconceptions.

If computer-assisted learning programs are used as part of an integrated statistics curriculum for psychology students, then as Shuell (1992) has suggested it is important to consider the capabilities of computers that provide advantages with regard to their use for instructional purposes. The design of a computer-assisted learning program must harness the instructional capabilities of computer technology. Computer technology can be used to provide the following facilities in a computer-assisted learning program:

(i) Direct manipulation and interaction. In the instructional process, WIMP interfaces (*w*indows, *i*cons, *m*ouse or *m*enus, *p*op-up or *p*ull-down menus or *p*ointing) (Preece, 1993, p. 82) provide opportunities for the student to select, highlight and manipulate objects. This is likely to mean that students find the program's interface

easy to use and they can therefore attend to the learning task at hand, rather than become bogged down in typing or data input (c.f. Milheim, 1995 - 1996).

(ii) Immediate feedback, which can be in the form of graphics, text or sound, can be provided to a learner in response to their actions at the human-computer interface. Feedback can be contingent on the learner's actions at the interface (Shuell, 1992). In other words, specific feedback, which is dependant on the student's response to say a particular learner activity or self-assessment question in the program, can be provided.

(iii) Multiple linked representations. Static graphics in the form of diagrams and pictures can be used in instructional texts, but in a computer-assisted learning program multiple representations can be dynamically linked. For example, at the interface, a learner could change the value of a statistic, such as the correlation coefficient and see the corresponding change in data displayed on a scatter plot.

(iv) Record keeping. A program can be designed to create and maintain a record of student responses and actions at the human-computer interface, the parts of the program to which they were exposed and the feedback they received (Shuell, 1992). Records in the form of student logs can be used to inform the learner of their progress, or can be used by the teacher and/or student so that they can monitor the learning process (Shuell, 1992). Student logs can be used by developers and evaluators of computer-assisted learning programs who might require formative data to improve the design of a teaching application (Jones et al, 1996). Indeed, research studies can also use student logs because they provide a source of data concerning students' interactions with the instructional system.

(v) Sound. Audio recordings can be played to the learner to avoid, for example, the over use of text in a program.

With the above considerations in mind, the first prototype of *Link* was designed to include learner activities that provide direct manipulation and interaction for a student and

use text, scatter plots and sound. For example, activity 3 provides correlation coefficients that can be selected and dragged to complete the activity and the feedback to activity 1 uses sound. Activity 2 provides two representations that are linked so that if a learner selects a particular correlation coefficient the program subsequently displays the corresponding scatter plot. The learner activities also provide informative feedback. The formative evaluation study provided qualitative data concerning students' responses to particular activities (chapter 8). Accordingly, the final version of *Link* was developed to include feedback to the learner activities that is contingent on a learner's actions at the interface (see chapter 9). In addition, the final version of *Link* was designed to create a student log of a user's interactions with the program. In the formative evaluation study, direct observation of students using the first and second prototypes of *Link* was used to record how students worked through the program and how they completed the different learner activities.

5.3 *Link*

Link was designed to be used as a revision program for psychology students who are in the second or final year of their undergraduate degree programmes. Empirical work, which was described in chapter 3, had identified misconceptions held by psychology students who had already completed courses in statistics that had covered correlations. The program was designed therefore for students to review their understanding of correlation.

Link was designed to cover linear correlation, and the concepts of positive correlation, negative correlation, zero correlation, the strength of correlations and correlation and causation. An introductory screen was designed to tell students that:

"In this package you will review your understanding of correlation. The aim of this package is to make sure that you have a clear idea about the different kinds of relationships that can be found between variables."

(Figure 5.1).

This screen was used because effective instructional materials should have some kind of advanced organiser or some other pre instructional introduction to prepare learners for the material that they will study (Bangert-Drowns & Kozma, 1989; Gagne, Briggs & Wager, 1988). In addition, the introductory screen allows the student to select a study so that they can review their understanding of correlation. The study TV violence and children's aggression (Eron et al, 1972) was adapted for the first prototype. The study by Eron et al (1972) was reported in a journal and it was therefore necessary to summarise and adapt the study so that an outline of the study could be prepared and used in the program (figure 5.2 and figure 5.3). Eron et al (1972) reported correlation coefficients that were obtained in the study between variables, such as a measure of a child's aggression and a measure of a child's preference for TV violence. The program uses correlations that are similar to the ones obtained in this study, but also provides correlations that are different from the ones obtained because a variety of different kinds of relationships were required for the learner activities in the program. For example, a negative correlation coefficient that was stronger than a positive correlation was needed for activity 2. Activity 1 and activity 2 use a table of data that was designed to display the correlation coefficients (figure 5.2).

Figure 5.1 The introductory screen of the first prototype

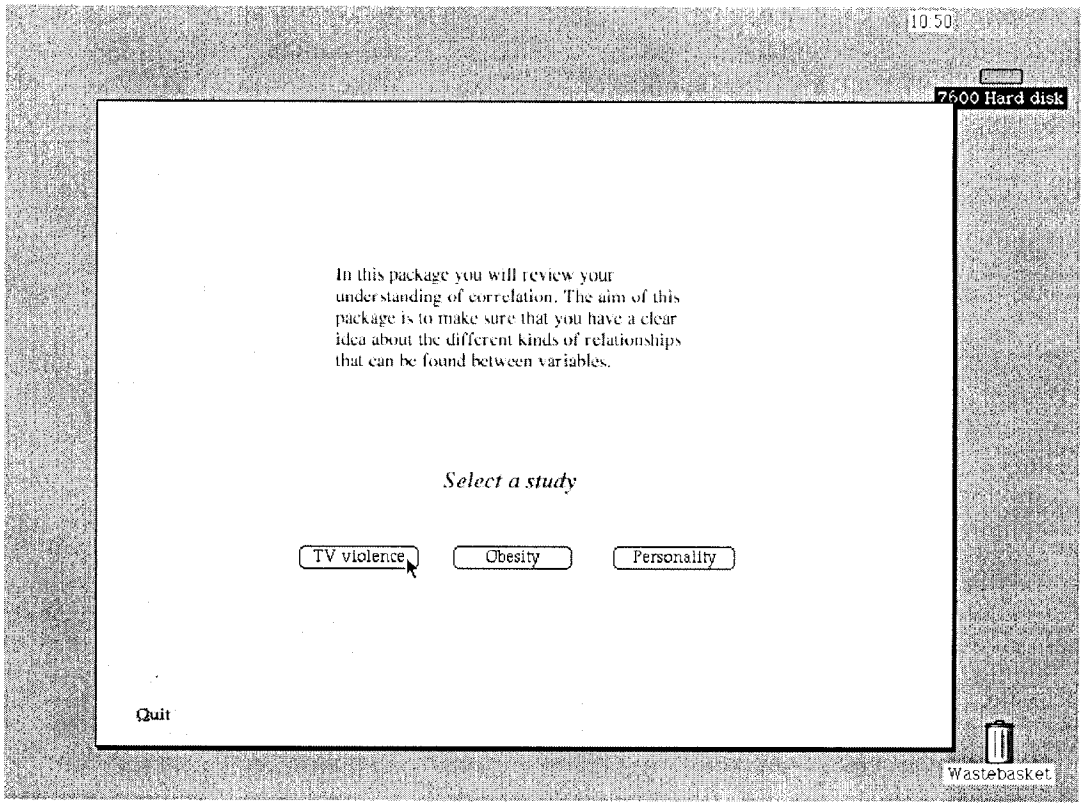


Figure 5.2 Introductory screen for the TV violence study

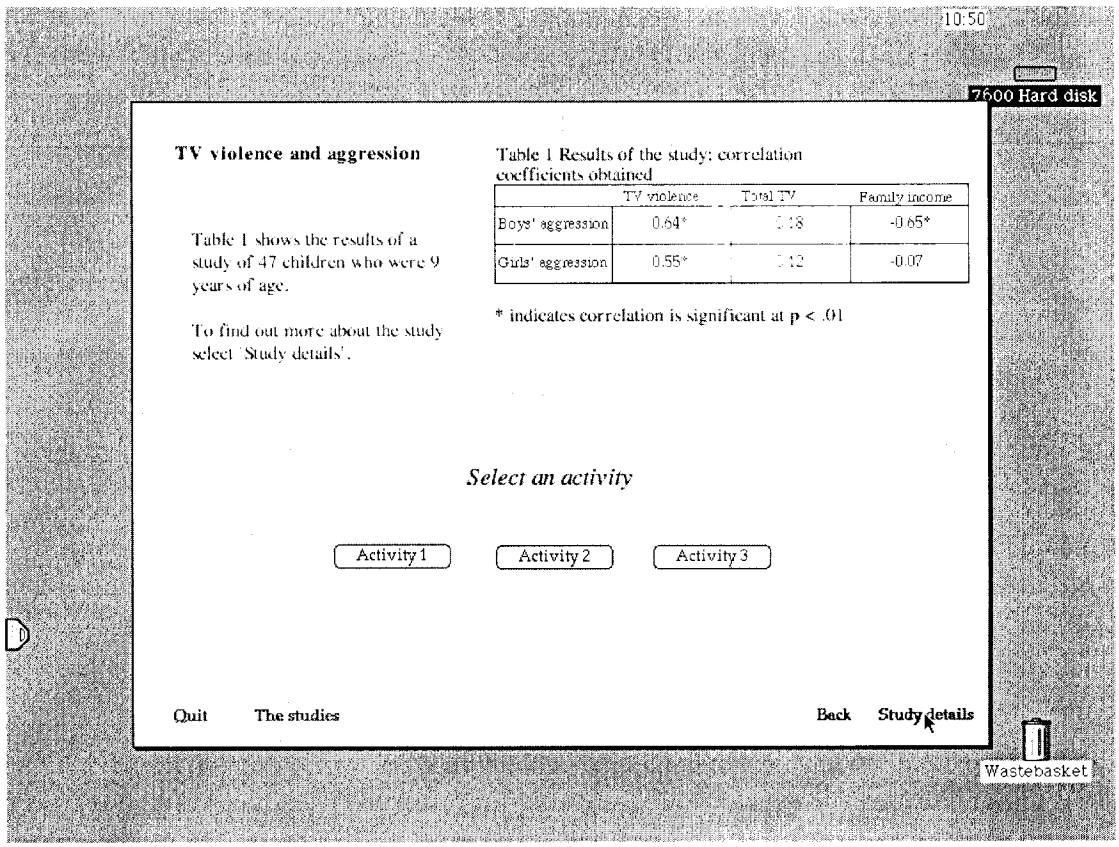
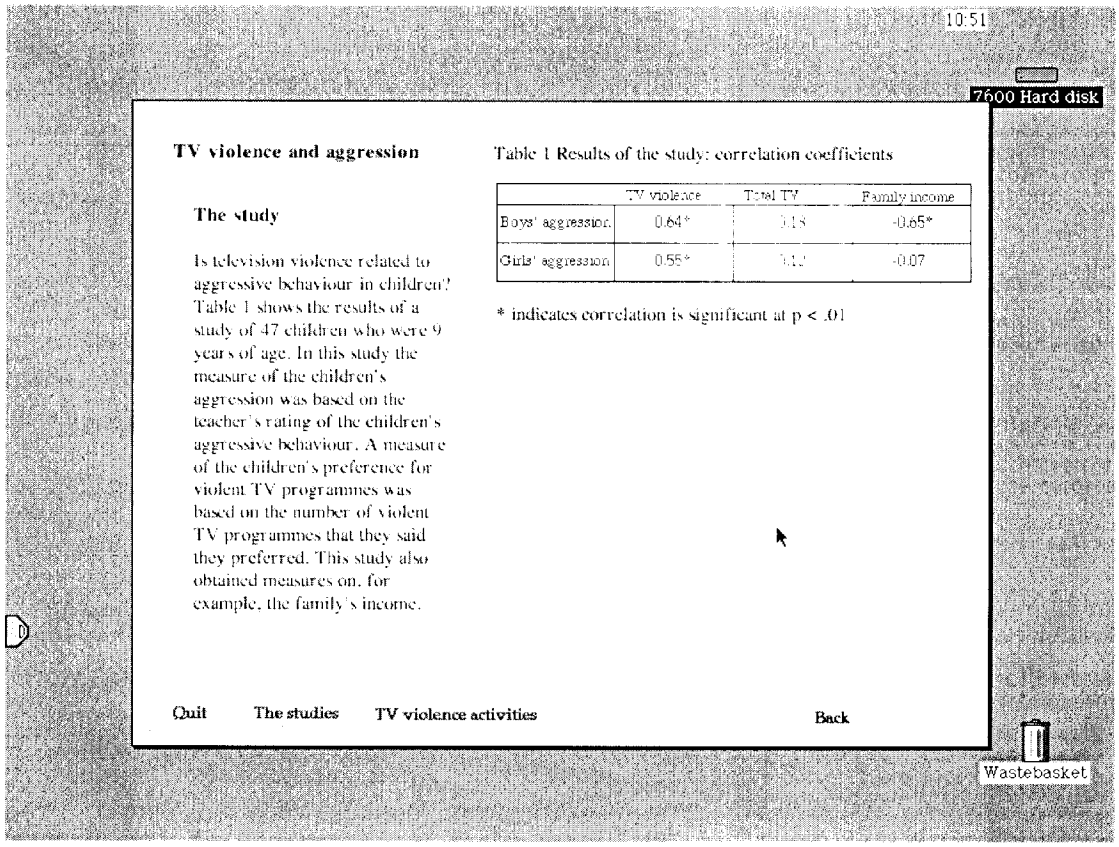


Figure 5.3 A screen that provides details of the TV violence study



It should be noted that although the coefficients used in the earlier versions of *Link* were based on coefficients from the TV violence study, the data sets themselves were not available. However, the final version of *Link* provides learner activities in the context of two different studies from the psychological research literature (Cohen & Java, 1995; Oates, 1998). Permission was given to use the data sets from these studies for *Link* and also to provide an outline of each study in the program (chapter 9).

For *Link*, three learner activities were designed to address specific misconceptions as follows:

- Activity 1. This was designed to address a causalistic conception of correlation by asking a student to interpret the meaning of a correlation coefficient. Feedback to this activity informs a student that four interpretations of a correlation are possible and asks the student to think of a third variable that might be responsible for the

relationship between the boys' aggression and TV violence. The strategy used in this activity is similar to a strategy described by Driver (1988): 'broadening the range of application of a conception' because a student's prior conception can be used as a resource which can be extended. A student might initially think that A is the cause of B, and this idea is built on by providing other possible interpretations of a correlation. For example, a student is to consider that B might be the cause of A or that an additional variable might be responsible for the obtained correlation.

- Activity 2. This was designed to address a unidirectional conception of correlation. Here, a student must decide which correlation from a set of coefficients represents the target scatter plot. The scatter plot represents a negative correlation. If a student selects any of correlation coefficients, the corresponding scatter plot is displayed along side the target scatter plot. The strategy used in this second activity is similar to a suggested teaching strategy: 'differentiation of a conception' (Driver, 1988). A student might hold a unidirectional conception of correlation, which means that their conception of correlation is global and ill-defined, and certain experiences are necessary to ensure that they differentiate their conception. Through activity 2 a student would not only see that a correlation can indicate a negative relationship between two variables, but they would also have the opportunity to view a negative correlation on a scatter plot compared with, for example, the pattern of a positive correlation and/or a pattern that indicates no relationship between two variables.

- Activity 3. This was designed to address the conception that a positive correlation is stronger than a negative correlation. A student can select and drag a set of correlation coefficients and arrange them in the appropriate order to indicate the weakest to strongest relationship. The activity is set up so that the strongest relationship is a negative correlation (-0.65), but a positive correlation is also provided (0.64). A student might position 0.64 as stronger than -0.65, but the program then provides feedback to the contrary. The strategy used in this activity is similar to the teaching strategy 'the construction of an alternative conception'

(Driver, 1988). In some cases, problems might arise if the students' ideas are used to shape formal models because students' prior conceptions (or misconceptions) are at odds with formal conceptions. Students' misconceptions are therefore acknowledged and the alternative formal model is put forward. With regard to this strategy it should be noted that Driver (1988) has suggested that students' prior conceptions should be acknowledged and also discussed and that students should have the opportunity to evaluate the formal model in relation to their prior conceptions. However, the strategy used in activity 3 does not *explicitly* encourage students to discuss their prior ideas (their misconceptions).

5.4 The development of *Link*

An iterative process of design was taken in the development of *Link*. This design process is described by the use of prototypes that animate some but not all the features of an intended system (Dix, Finlay, Abowd & Beale, 1993). More specifically, the approach of evolutionary prototyping was used where a prototype is not discarded, but serves as the basis for next design iteration (Dix et al, 1993).

The authoring tool Macromedia Director 5.0 for the Macintosh was used to develop *Link* because elements, such as graphics, sounds, buttons and text can be easily incorporated in a program. Director provides a prototyping tool where screens that include text and graphics can be easily produced. In addition, Director's scripting language, Lingo can be used to add an interactive dimension to a program (Persidsky, 1996). A Power Macintosh 7600/120 was used as a development platform. *Link* was developed to run on a Macintosh platform because it was known that students who would take part in the formative and summative evaluation studies of the program were at institutions that had Macintosh computers in the laboratories available for student use.

The development of *Link* involved two phases of a formative study with target users of the program and an expert evaluation. The first phase of the formative study

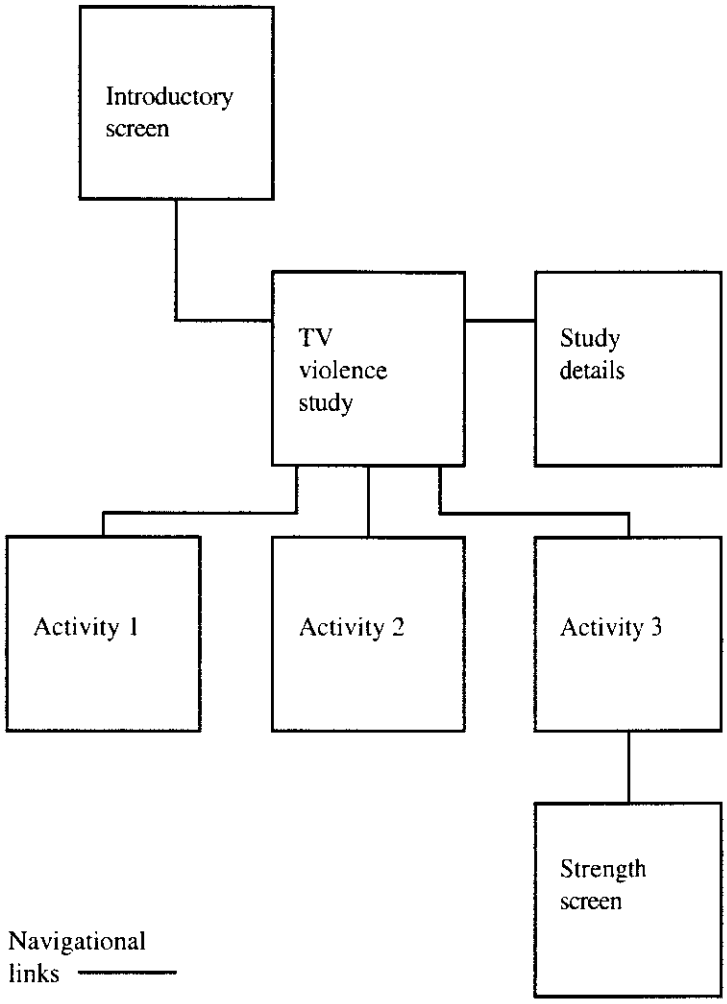
informed the development of an improved second prototype, which was subsequently used in the second phase of the formative study and was also examined by specialists.

5.5 Program implementation

Figure 5.4 is a schema that describes the organisation of the first prototype. The program is comprised of six movies created in Director. Each movie is represented by a screen as depicted in figure 5.4. For example, the introductory screen for the TV violence study presents text and a table of data, but also provides buttons that a user can select to invoke one of the screens that presents a learner activity (figure 5.2). By using simple Lingo scripts, the cast members of a movie can be used to provide interactivity. For instance, the button 'activity 1' contains the script:

```
on mouseUp
  play movie "activity 1"
end
```

Figure 5.4 Schema that describes the organisation of the first prototype



So, if a user selects this button, the movie called 'activity 1' is played and the screen that provides this activity is displayed to a user. Interactivity was also added to the three learner activities by using the play command. If a user selects the button 'Done' when they have answered a learner activity, feedback is provided at the interface. This was achieved by using a script that specifies that a particular frame of a movie is to be played when the button is clicked. The frame that represents a screen of the movie, which presents feedback, is therefore displayed. Appendix C provides the Lingo scripts that were used for *Link*.

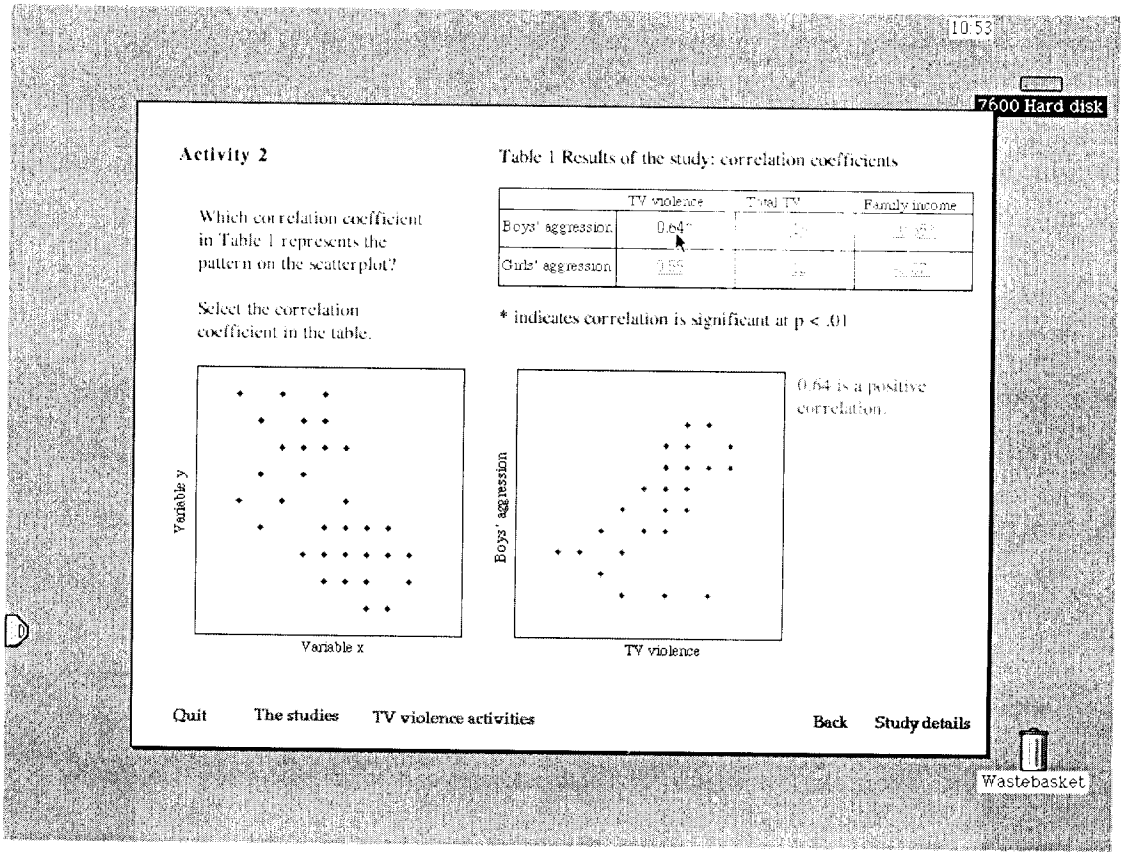
It has been suggested that simply stating whether an answer to a question or activity is correct or incorrect is not sufficient for effective learning, but that feedback from a

computer-assisted learning program should be relevant to a learner and tied to a specific answer provided by a learner (Milheim, 1995 - 1996). However, research on student modelling has shown that even though different types of student models can be devised, student modelling is a difficult problem with no easy solution (Ohlsson, 1993; Wenger, 1987). For example, it is extremely difficult to anticipate all conceivable errors that a student might make and it is also difficult to determine how relevant information about a student's knowledge state is acquired by the system so that appropriate feedback can be provided.

The outcome of the expert evaluation of the second prototype of *Link* suggested that feedback to an activity should be contingent on a user's response to that activity (chapter 8). By using Lingo, feedback can be provided that is conditional on a user's responses at the interface. Iterative if...then...else structures can be used to test when a condition exists and the program can respond accordingly. In this way, the final version of *Link* provides specific feedback to learner activities, which will be considered in chapter 9.

Figure 5.4 depicts some of the navigational links that are provided by the program. For instance, by selecting the appropriate buttons, a learner can move from the introductory screen of *Link* to the TV violence screen and from here select activity 2. With regard to the links shown in figure 5.4, there are, however, other navigational routes that can be taken by a learner. There are buttons at the bottom of the screen in the first prototype that allow a user to move from one screen to another. For example, from activity 2 a user may choose to find out more about the study and select the button 'Study details', which will invoke the screen that provides details about the TV violence study (figure 5.5 and figure 5.3). In the movies, transition effects, such as dissolve were set to occur when a user moved from one screen to another or when a scene in a movie changed to, for example, provide feedback to a user. Without transitions, Director scenes simply cut abruptly from one scene to the next and therefore often create a jarring effect for a user (Persidsky, 1996).

Figure 5.5 Activity 2



To create the correlation coefficients and scatter plots that were required in the program, *SPSS* was used. A variety of coefficients were required for the three activities including positive correlations, a strong negative correlation, (which was stronger than the positive correlations), and a correlation that was very near zero to represent no relationship. Sets of data consisting of whole numbers for variables, such as TV violence, were entered into an *SPSS* spreadsheet. By using *SPSS*, the required variety of Pearson correlation coefficients were obtained by experimenting with different sets of data and changing entries in the sets if necessary. The application CA-Cricket Graph III was used to generate a table of data that contained the names of variables and the correlation coefficients, and the scatter plots that corresponded to the correlation coefficients obtained between variables. This data table and the scatter plots were imported to the program as required.

Activity 1 uses sound: audio feedback is provided to this activity. To achieve this, the voice of the researcher reading small sections of text for the feedback was simply recorded by using a microphone attached to the authoring platform. Sound Edit 16 was used to record the chunks of text that were saved as AIFF sound files and were used in activity 1.

An application program can be easily produced with Director by creating a projector movie which is a play-only version of a movie. To start the program, the projector movie that represents the introductory screen of *Link* can be run. To run, *Link* requires a 68020 Macintosh, running System 7, at least 4 megabytes of RAM and a 640 x 480 monitor that is set at 256 colours.

5.6 Using *Link*

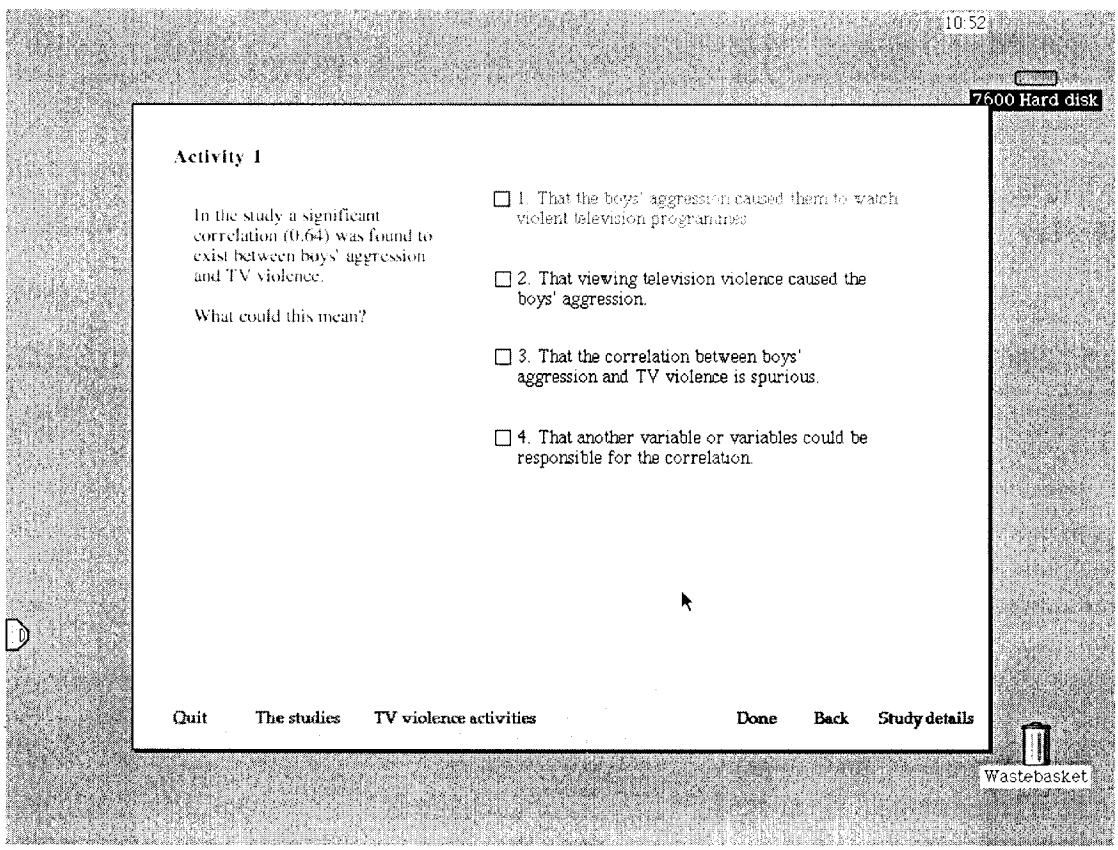
On starting the program, a user is provided with the introductory screen from which they can select the button 'TV violence' (figure 5.1). By selecting this button, a screen that outlines the TV violence study and also provides buttons for the three learner activities is presented to a user (figure 5.2). A user can find out more about the study by selecting the button 'Study details', which takes them to a screen providing further details about the study (figure 5.3). Here, a user can then select the button 'TV violence activities' and they are taken back to one of the introductory screens from where they can select 'Activity 1'.

On activity 1 a student is asked the possible meaning of a significant correlation that was found to exist between boys' aggression and TV violence. A student can select a maximum of four options that provide possible interpretations of this correlation. When a student has checked one or more answers to the question in activity 1, they can then select the button 'Done' that is provided at the bottom of the screen. This invokes audio feedback that informs the student that all four interpretations are possible:

"For a correlation, all four interpretations are possible. It could be that the boys' aggression caused them to watch the violent television programmes, or that viewing television violence caused the boys' aggression. It could be that the correlation between the boys' aggression and the TV violence is spurious. This means that the finding was simply due to sampling error. A third variable could be responsible for the obtained correlation. What kind of third variable could be responsible for the relationship that was found to exist between boys' aggression and TV violence?" (Audio feedback to activity 1).

When the audio feedback is played, the relevant options are highlighted in turn in accord with the possible interpretations of a correlation that are described (figure 5.6).

Figure 5.6 Activity 1. Options are highlighted in accord with relevant audio feedback



At the end of this audio feedback a student is asked to think of a third variable, such as

parenting style or parental aggression that might be responsible for the relationship between the boys' aggression and TV violence.

Activity 2 was designed to address a unidirectional conception of correlation and makes use of the study data in the form of correlation coefficients presented in a table (figure 5.5). A student is not told that the scatter plot that is presented on the screen represents a negative correlation, rather, they are asked to decide which correlation in the table represents the pattern on this scatter plot. If a student were to select 0.64 in the table, a scatter plot showing this positive correlation is displayed alongside the target scatter plot and the student is provided with feedback that states "0.64 is a positive correlation" (figure 5.5). A student can select any of the correlations in the table and feedback and the appropriate scatter plot will be displayed.

The third activity in this first prototype of *Link* was designed to address the conception in which a student thinks a positive correlation of say, 0.80 is stronger than a negative correlation of say, -0.90. This activity also makes use of the table of data, but here a student can drag the correlation coefficients from the table to arrange them in an appropriate sequence. This activity involves a student selecting the coefficients in turn in the table and dragging them to arrange them in an order from that which represents no relationship to that which indicates the strongest relationship between two variables (figure 5.7). When a student completes this task, they select the button 'Done', and feedback is provided at the interface so that they can compare their arrangement with the correct one that is shown (figure 5.7). The feedback to this activity also tells a user to select the button 'Strength' which, if clicked, presents an additional screen for this third activity. This screen provides a scale that runs from -1 to 0 to 1. On this scale, a student can select various points, which indicate to a student how both positive and negative correlations are strong if they are relatively near to a coefficient of -1 or 1 (figure 5.8).

Figure 5.7 Activity 3

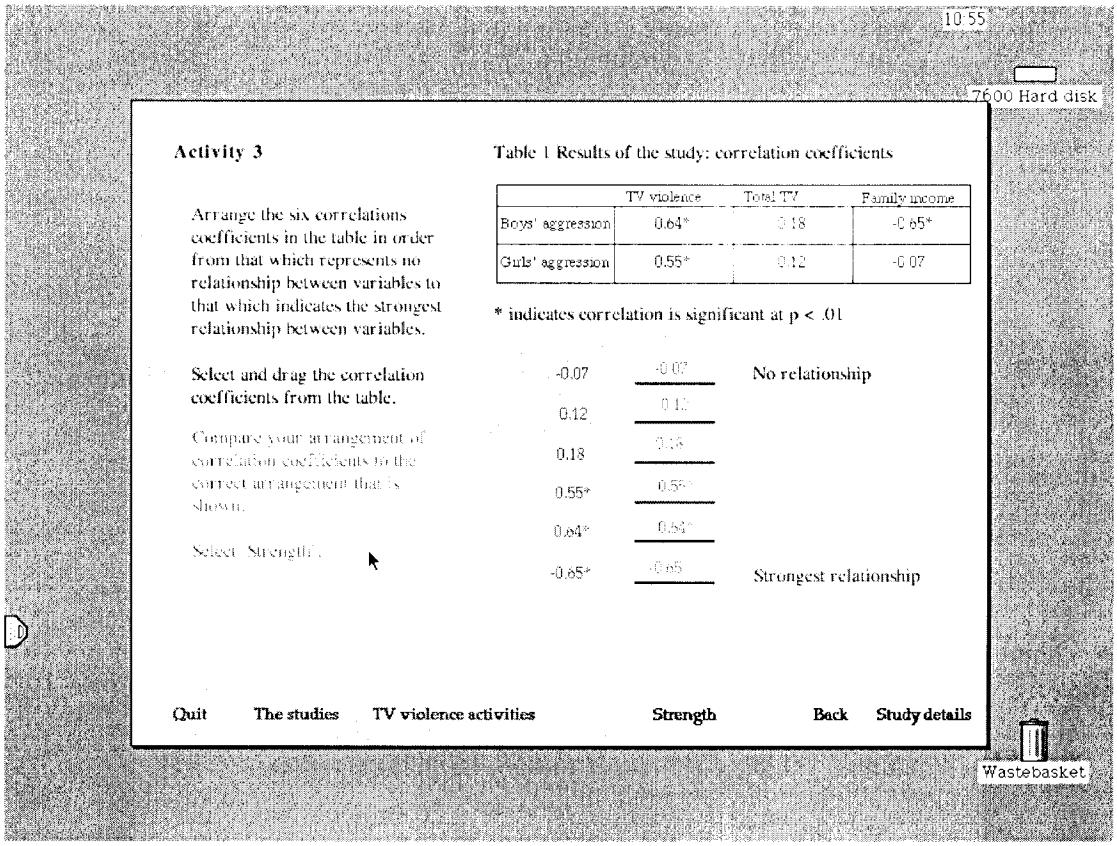
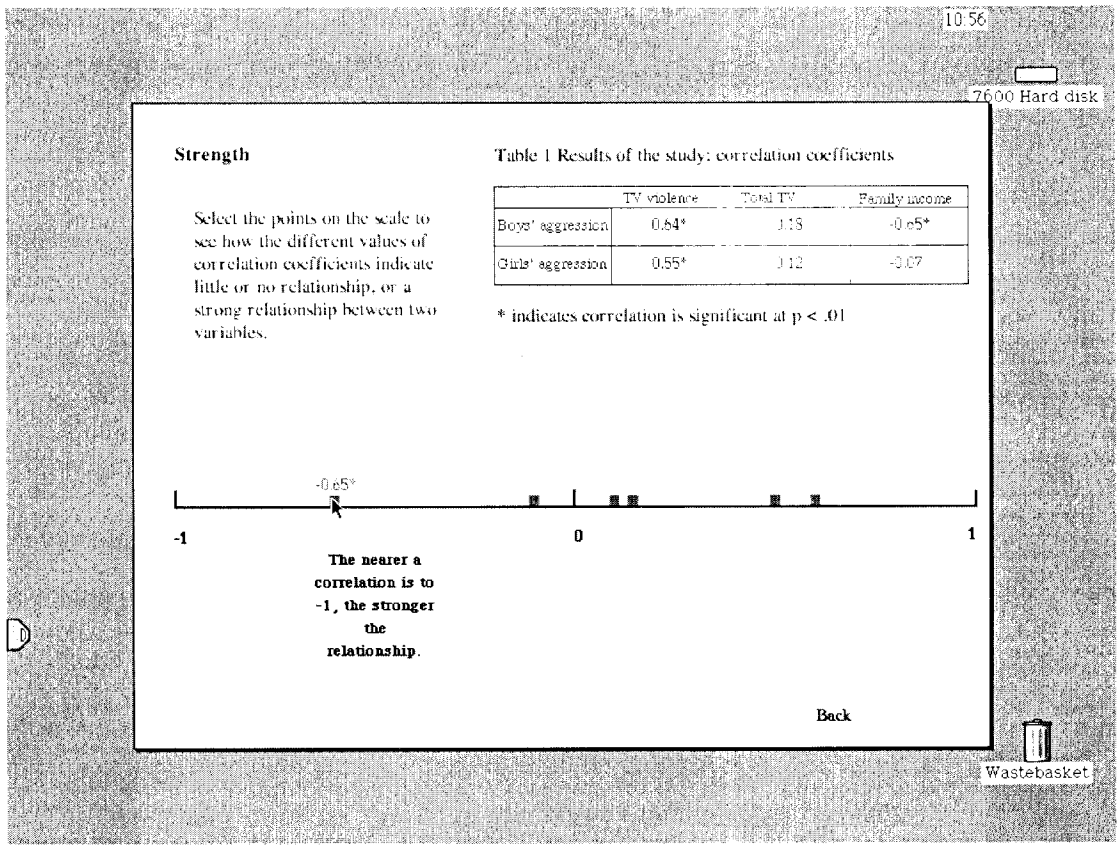


Figure 5.8 An additional screen for activity 3



5.7 Summary

This chapter described the design and development of the first prototype of *Link*. *Link* was designed to be used as a revision tool by psychology students who have already studied correlations. It was also developed to investigate whether particular kinds of learner activities in a computer-assisted learning program contribute to students' understanding of correlation. The design of the prototype was informed by general principles of learning, empirical work that has identified students' misconceptions pertaining to correlation and a review of research and developments in the field of computer-assisted learning. Using Macromedia Director, a prototyping approach was taken in the development of the program involving a formative evaluation and an expert evaluation of prototypes. The first prototype provides three learner activities in the context of a study that concerns TV violence and children's aggression. Each of the learner activities in the program were designed to address particular misconceptions.

Chapter 6

Methodology for the evaluation of computer-assisted learning programs

6.1 Introduction

This chapter describes the methodology that was employed in the evaluation studies of *Link*. It will be argued that the development of computer-assisted learning materials should involve formative evaluation (Laurillard, 1993). Furthermore, empirical work should be conducted to investigate a program's effectiveness in terms of students' learning. Summative evaluation can be conducted to investigate not only whether using a program contributes to students' understanding of a particular subject area, but also to find out about how students learn from a program.

As part of program development, a formative evaluation of *Link* was carried out, which was followed by an expert evaluation of the program. In the light of the findings from the formative study and the expert evaluation, a final version of *Link* was developed and empirical work in the form of a summative evaluation was conducted. This chapter considers pertinent issues in the evaluation of computer-assisted learning programs and then outlines the evaluative framework that was employed for the evaluation studies. The methods and instruments that were used to gather both quantitative and qualitative data in the formative and summative studies are then described.

6.2 The evaluation of computer-assisted learning programs

Formative evaluation can be used as part of the development of a computer-assisted learning program. This kind of evaluation should be conducted to ensure that students find the program easy to use and can provide information for improvement of the learning materials (Laurillard, 1993). The findings of a formative evaluation can therefore inform further program development. In addition, formative evaluation of educational materials

with the target students provides the opportunity for an intensive look at how students learn through educational media (Laurillard, 1993).

It is clear that computer-assisted learning programs should be pleasant and easy to use so that the student can attend to the learning task at hand (Bangert-Drowns & Kozma, 1989). If a program has a poorly designed human-computer interface where, for example, the text is unclear or the function of a button is ambiguous, then the student might become confused or be distracted from the subject concepts or learning activities that are provided.

To design an effective human-computer interface, Hix and Hartson (1993) have recommended that formative evaluation is carried out early on in the development of a program so that usability problems can be uncovered when there is still sufficient time for modifications to be made to the design of the program. It is recommended that for the purposes of improving the human-computer interface only a small group of target users should be recruited for a formative evaluation study (Hix & Hartson, 1993; Monk, Wright, Haber & Davenport, 1993). After five or six users, the participants tend to stop finding novel problems and usually reiterate ones that have already be discovered by prior participants (Hix & Hartson, 1993). Similarly, Monk and his colleagues (1993) have concluded that five is the maximum number of participants required in the evaluation of a prototype. Hix and Hartson (1993) have suggested that a variety of data can be collected in a formative evaluation of a program's human-computer interface. The kind of data that is gathered will of course depend on the type of system that is being evaluated. They have recommended that qualitative data, such as a list of the problems that users face when using a particular interface should be generated because this might result in suggestions for modifications to improve the design of the human-computer interface. The generation of such qualitative data requires the development of tasks for participants to perform while they use a program. These should be representative tasks that users would be expected to carry out with the developed program.

Both Hix and Hartson (1993) and Monk et al (1993) have provided recommendations and techniques for the formative evaluation of a program's human-computer interface. Participants should be recruited from the target population of users of the program. In field testing, the program is brought to the participants and the current prototype is set up in the normal working environment in which the users are expected to use the program (Hix & Hartson, 1993). Participants are of course informed about the procedure for evaluation by using "introductory instructional remarks" (Hix & Hartson, 1993, p. 297). It is suggested that participants are observed while they work with the program by completing preplanned representative tasks. The user is asked to talk aloud while they complete the tasks, providing qualitative data, such as critical incidents that might occur when they are using the program. Monk and his colleagues (1993) have provided valuable questions that can be used to ensure the participant continues to give a running commentary of what they are doing and thinking while they use the program. After the participants have completed the tasks, they can be asked to complete a questionnaire that is designed to elicit subjective comments about the prototype.

In the light of the above, phase one of the formative evaluation of *Link* was carried out. As shall be detailed in chapter 7, in this first phase, six students tried out the first prototype of *Link* in a computer laboratory at their university. This phase was conducted to assess the quality of the prototype's human-computer interface and to identify usability problems that might exist. The findings of this phase were used to modify the first prototype and to develop an improved program that was used in the second phase of the formative study.

There have been recommendations in the form of frameworks for the evaluation of computer-assisted learning programs (Draper, Brown, Henderson & McAteer, 1996; Draper, Brown, Edgerton, Henderson, McAteer, Smith & Watt, 1994; Jones, Scanlon, Tosunoglu, Ross, Butcher, Murphy & Greenberg, 1996). It has been suggested that the evaluation of computer-assisted learning should be empirical and that students themselves

participate in the evaluation of a program (Jones et al, 1996; Reiser & Kegelmann, 1994). Draper and his associates (1996) also argue that the approach to the evaluation of education software should be empirical and should not be based solely on, for example, an expert opinion of the product in question.

It has been recommended that a variety of sources of information are used in an evaluation study in which qualitative and quantitative data is generated and collected (Jones et al, 1996). Jones et al (1996) have highlighted that there is not only a need to include learners in the process of evaluation, but what students have learnt as a result of using a program should be assessed where it is possible. In addition, students must be observed while they use the program and asked to give their opinions about the program in question. Valuable data about the quality of the program that the students worked with is therefore provided. In a similar vein, the approach advocated by Draper et al (1996) has suggested that a variety of instruments, such as knowledge quizzes, and direct observations of students working with the program should be used in an evaluation study. Knowledge quizzes are related to particular learning objectives and are designed to provide a quantitative measure of whether a students has, for example, acquired a particular concept (Draper et al, 1996). Observing students working with the program will of course provide qualitative data concerning particular problems that they might encounter with the software. In the evaluation of educational software, Zahner and her colleagues have also stressed that it is critical to obtain what they term performance data so that it can be determined if students achieve the learning objectives that the software is designed to teach (Zahner, Reiser, Dick & Gill, 1992).

Jones et al (1996) have described a principled approach to the evaluation of computer-assisted learning programs. The evaluative framework they have outlined provides recommendations to the kinds of data that should be generated and gathered in the evaluation of a program. This framework, which is outlined in the next section, was employed in the evaluation studies that are described in this thesis for the following

reasons. First, the framework was based on a review of the pertinent literature that concerns the evaluation of computer-assisted learning. It has also made recommendations that have been proposed by related and similar frameworks (e.g., Draper et al, 1996). In addition, the approach provides a framework that can be adapted to organise the different sources of data that are required in an evaluation study. The evaluative framework was used in the second phase of the formative study and it was employed for the summative evaluation study of *Link* that is described in chapter 9.

In the research described in this thesis, a formative evaluation study and an expert evaluation were carried out as part of *Link's* design and development. The formative study was also carried out to pilot the questionnaires, pre and post tests and related instruments that were used in the summative evaluation study of *Link*.

The pre-test-post-test control group design, which is appropriate for investigating the effects of educational innovations, is commonly used in educational research (Dugard & Todman, 1995). In this kind of design, participants are assigned to a treatment or control condition and scored on a test both before and after taking part in one of the conditions. The fundamental features of this design are not changed by using additional treatment groups (Dugard & Todman, 1995). The evaluation studies of *Stat Lady*, which were outlined in the previous chapter, exemplify this kind of design and have looked at the efficacy of learning from a computer program (Shute et al, 1996; Shute & Gawlick-Grendell, 1994). Both of these evaluation studies used two treatment groups where learning from *Stat Lady* was compared to learning from a series of statistics lectures (Shute et al, 1996) and a paper-based Workbook covering the same curriculum (Shute & Gawlick-Grendell, 1994).

However, concerns that relate to evaluation studies of computer-assisted learning programs that use pre-test-post-test control group designs have been raised (Draper et al, 1996; Hawkins et al, 1992). Draper and his colleagues (1996) have argued that:

"It is not sensible to design experiments to show whether CAL [computer-assisted learning] is better than lectures, any more than whether textbooks are good for learning: it all depends on the particular book, lecture, or piece of CAL" (ibid., p. 27).

This is because learning is determined by a variety of factors that will inevitably vary across situations (Draper et al, 1996). The results of an experiment that was designed to test whether a computer-assisted learning program contributed to students' learning a particular element of subject matter, as opposed to learning from say, a text, cannot be generalised to another context and cannot therefore predict the efficacy of learning from the computer program in another situation (Draper et al, 1996). Similarly, Hawkins et al (1992) have pointed out that it is not really possible to define the precise features of an instructional method and evaluate the learning outcomes in a way that is appropriate for each instructional method. If two instructional methods are therefore compared, it is extremely difficult to determine the factors that might have produced apparently different learning outcomes (Hawkins et al, 1992).

The summative evaluation of a computer-assisted learning program is, however, essential: it can provide important empirical findings concerning the process of student learning and can indicate which features of a program can contribute to students' understanding of a particular area. Clearly, simple pre-test-post-test control group designs that only collect data concerning learning outcomes are insufficient because they simply tell us whether students have learnt from a program or not. Yet summative evaluation studies can be designed so that data related to learning outcomes *and* the learning process can be gathered.

With regard to an evaluation study that is designed to focus on a computer-assisted learning program as the educational innovation, it is important to include a control group because otherwise learning gains, which might be measured by students' performance on the pre and post tests, could simply be attributed to a practice effect. In

other words, a repeated measures design means that participants gain practice at answering particular kinds of questions and therefore tend to score higher on a post test. In the evaluations of *Stat Lady*, the students in the control groups were only administered the pre and post tests at the same time as those students in the treatment groups (Shute et al, 1996; Shute & Gawlick-Grendell, 1994). Students who take part in a control group should, however, complete an activity that is comparable in terms of cognitive effort to the activity in the treatment group(s). Therefore, the summative evaluation study of *Link* used a basic control group that involved students working through a section of a computer-assisted learning program that did not cover correlation.

An additional form of control was necessary in the empirical evaluation of *Link* because it needed to be ensured that the learning gains that might be made by students who used the program could not be attributed to the fact that students simply cover a topic that could be presented by paper-based instructional materials. As well as the basic control group, the summative evaluation study therefore used an instructional control group that involved students completing paper-based instructional materials covering correlation.

One of the main interests in the evaluation of *Link* was to investigate the learner activities in the program. The summative evaluation study of *Link* was set up to investigate the learner activities in *Link* and whether they addressed students' misconceptions that concern correlation.

6.3 A framework for the evaluation of *Link*

The framework described by Jones et al (1996) is comprised of three main dimensions: context, interactions and outcomes. It has been used in the evaluation of computer-assisted learning programs that have been developed for courseware used by the Open University. In such cases, the context of the courseware must be considered. For example, the developers of computer-based learning materials should be involved in the

design of an evaluation study because the rationale and main objectives of the courseware would need to be determined (Jones et al, 1996). For the purposes of phase two of the formative evaluation and the summative evaluation study of *Link*, the framework was adapted to primarily focus on two of the dimensions: interactions and outcomes as shown in table 6.1.

Table 6.1 Framework for evaluation. (Adapted from Jones et al, 1996, p. 12).

	Interactions	**Outcomes**
Methods	Observation. Think-aloud. Audio recording. Student logs	Questionnaires. Tests.
Data	Records of student interactions and think-aloud.	Measures of learning.

6.3.1 Interactions

In the evaluation of a program, students should be observed and asked to think aloud while they work with the software. This provides qualitative data about the learning process. In the case of a formative evaluation, this also provides qualitative data about whether students experience particular difficulties with the learner activities that are provided at the human-computer interface. Think-aloud can be audio recorded for later analysis. The software can be set up to create and maintain student logs which can provide a record of students' interactions with the application.

6.3.2 Outcomes

It is important to investigate whether students have learnt the subject matter that the program is designed to impart or whether students' achievements have met the requirements of the learning objectives of the program. With regard to the particular learning objectives of the program, learning outcomes can and should be measured by, for example, carefully designed tests that are completed by the learner both before and after they use the program to see if there is a change in the students' performance on such tests.

6.4 Formative evaluation

The formative evaluation study involved students from Buckingham University and an Open University residential school who were studying psychology. As noted above, the first phase of the formative evaluation study was designed to assess the usability of the first prototype of *Link*. The second phase of the study focused on whether an improved version of *Link* contributed to students' general understanding of correlation, but was also designed to test the use of questionnaires, pre and post tests and instruments that would be used in the summative evaluation study.

The formative evaluation study used a questionnaire for participant details, participant instructions, a set of tasks for the evaluation session, a form to record data and a program evaluation questionnaire. The second phase of the study also used tests in correlation that were completed before and after students used *Link*. These materials are detailed in chapter 7 and chapter 8 which describe the formative evaluation study. Phase one of the formative study involved students working individually with a prototype of *Link*: they completed set tasks and were observed and asked to think aloud while they used the program. A set of questions was also used in the evaluation session to make sure that the participants continued to give a clear commentary of what they were doing and thinking while they used the program. After students had used the program they

completed an evaluation questionnaire that was designed for participants' opinions about the program. Phase two of the formative evaluation followed the same format as the first phase: students were set tasks to complete while they worked with the program and were asked to think aloud while they did this. This phase also looked at students' learning and in order to assess learning outcomes, equivalent tests in correlation were developed that were designed to provide an assessment of a student's understanding of correlation and to identify particular misconceptions that a student held. The questions in these tests were used in the investigation that was outlined in chapter 3, and were adapted for the formative study. The questions for the tests were examined for accuracy and clarity by two subject specialists and were modified accordingly for the developed tests. The evaluative framework in table 6.2 was used to organise the different kinds of data that were generated and collected in the second phase of the formative study.

Table 6.2 Framework for the second phase of formative evaluation

	Interactions	**Outcomes**
Methods	Students completed tasks and were observed while they worked with program. Students were asked to think aloud while they completed the tasks. A record of observations and audio recording of students' think-aloud were made.	Questionnaires designed to elicit students' opinions about the program were used. Tests were completed by the students before and after they used the program.
Data	**Qualitative** Records of student interactions and think-aloud. This record included details of how students answered the learner activities.	**Qualitative** Students' opinions about the program. **Quantitative** One-tailed related t-test was used to see if there was a significant difference in the mean scores of the pre and post tests. **Qualitative** The students' answers to questions on the tests were examined to identify particular misconceptions. **Quantitative** McNemar tests were used to see if students' conceptions had changed from the pre test to the post test.

Three different subject specialists took part in the expert evaluation of *Link*. They were asked to evaluate the program by working through it and by completing the evaluation questionnaire that was used in the formative study. Chapter 8 describes the findings and implications of this expert evaluation with reference to the development of *Link*.

6.5 Summative evaluation

The overall aim of the summative evaluation was to investigate whether the final version of *Link* contributed to students' understanding of correlation and the more specific aim of the study was to investigate the use of learner activities in the program and whether they affected students' misconceptions in correlation.

The summative evaluation study that is described in chapter 9 was quasi-experimental (Clark-Carter, 1997) and used a pre-test-post-test control group design (Dugard & Todman, 1995). As described above, the focus of this evaluation was whether *Link* contributed to students' understanding of correlation and the study used two control groups. This design is detailed in chapter 9.

The formative evaluation study served to test the use of the techniques, questionnaires and tests for the summative study. The participant profile, and the tests in correlation were modified for use in the summative evaluation study. In this study, learning outcomes were assessed by students' scores on the pre and post tests in correlation. ANCOVA was applied in the analysis of the pre and post test scores for the three different groups, which is an appropriate and informative technique for data from a pre-test-post-test control group design (Dugard & Todman, 1995). Qualitative data concerning students' interactions with the program were recorded by means of student logs that were created by the program. These logs provided a record of the students' actions at the human-computer interface (e.g., which buttons in the program were selected and which screens where invoked), students' responses to the learner activities

and the feedback that was received on the activities. More detailed data concerning the process of learning while students used *Link* was required and so, seven of the students who participated in the study were assigned to act as case studies. These students were observed and ask to think aloud while they used *Link*.

Two equivalent tests in correlation, which were adapted from the tests that were used in the formative study, were devised for the summative study. As for the formative study, the questions on these tests were developed to identify students' misconceptions concerning correlation. This meant that non-parametric statistical techniques could be used to see whether students' answers to particular questions changed after they had used *Link*.

Table 6.3 shows the application of the evaluative framework to the summative evaluation.

Table 6.3 Framework for the summative evaluation

	Interactions	**Outcomes**
Methods	Students completed tasks while they worked with the program. **Case studies** Students were observed and asked to think aloud while they completed the tasks. A record of observations and audio recording of students' think-aloud were made.	A pre-test-post-test control group design was used. *Link* group: tests were completed by the students before and after they used the program. Instructional control group: tests were completed by the students before and after they worked through paper-based instructional materials that covered correlation. Basic control group: tests were completed by the students before and after they used a program that did not cover correlation.
Data	**Qualitative** Student logs. These logs provided a student's route through the program and a record of how a student answered each learner activity. **Case studies** Records of student interactions and think-aloud.	**Quantitative** ANCOVA on post test scores, with pre test scores as a covariate **Qualitative** The students' answers to particular questions were examined to identify misconceptions. **Quantitative** McNemar tests were used to see if students' conceptions had changed from the pre test to the post test.

6.6 Summary

This chapter has described a methodology for the evaluation of a computer-assisted learning program. It was emphasised that the evaluation of a learning program should be empirical, involving students in the evaluation process and by collecting both quantitative and qualitative data. An evaluative framework was outlined that was used in both the formative and summative evaluation studies of *Link*. This framework was adapted from the research literature (Jones et al, 1996) and is an approach in which information from a variety of sources is collected in the process of evaluation. This framework was used so that data concerning learning outcomes, as measured by students' performance on pre and post tests, and data relating to the learning process, such as records of students' interactions with the program, were collected in the evaluation studies.

The formative and expert evaluation of *Link* were conducted as part of program development. The final version of *Link* was evaluated by means of a summative study that used a pre-test-post-test control group design to investigate whether *Link* contributed to students' understanding of correlation. The following three chapters describe the evaluation of the program.

Chapter 7

Formative evaluation: phase one

7.1 Introduction

This chapter describes the first phase of the formative evaluation of *Link*. This phase was conducted to assess the quality of the first prototype's human-computer interface and to identify usability problems that might exist. This formative testing of the program provided qualitative data that was used to inform the design of an improved prototype that was evaluated in the second phase of the formative study.

7.2 Method

7.2.1 Design

A formative evaluation of the first prototype of *Link* was conducted with target users in a computer laboratory at their university.

7.2.2 Participants

Six female students who were studying at Buckingham University, who were in the final year of their B.Sc. degree programme in Psychology with English Studies took part in the study. The mean age of this group of students was 23 years (mean = 22.83, S.D. = 2.99, minimum 20, maximum = 27). As part of their degree programme, all of the students had completed two courses in statistics that included a computing component: Quantitative Methods for Psychologists and Statistical Analysis for Psychologists. During these courses, the students used the data analysis software, Minitab. All of the six students had used an IBM PC-compatible before and one student had used both a PC and an Apple Macintosh. Five of the students reported that they had used Microsoft Word and all six of the students had used computers for at least a year and a half (median = 3 [2.75] years, minimum = 1.5, maximum = 5). Three of the students reported that they used

computers once a week, two students used computers every 2 to 3 days, and one of the students used a computer every day. To ensure anonymity an identifier (e.g., B1) is used for the different participants.

7.2.3 Materials

Hardware and software

The first prototype of *Link* was stored on an external hard drive for the Apple Macintosh. This was attached to a Apple Macintosh LC 475 so that the program could be run from the Macintosh desktop. The monitor was set at the required 640 x 480 resolution and at 256 colours.

Participant profile

This questionnaire was completed by the students to collect data concerning, for example, the participants' gender, age, qualifications, degree and year of study, the subject courses they had taken, and their computing experience. Part of this questionnaire was based on a sample Computer Experience Questionnaire (Draper et al, 1994).

Instructions for program evaluation

This outlined the purpose of the session and told the participant what would happen during the session. These instructions also told the user that they were to be observed while they used the program. They were informed that they were to complete a set of tasks and were asked to think aloud while they did this. The instructions were adapted from Hix and Hartson (1993, p. 299).

Task sheet

This contained the following five tasks that were carried out by the participants:

- Select the study 'TV violence'.
- Find out more about the study 'TV violence'.
- Do 'activity 1'.
- Do 'activity 2'.
- Do 'activity 3'.

Form for data

This provided a schedule to collect qualitative data, such as participants' actions and comments in relation to the specified tasks.

Evaluator's question sheet.

The questions on this sheet were used to make sure that the participant continued to provide a running commentary of what they were doing and thinking while they used the program. These questions were taken from Monk et al (1993, p. 83):

How do we do that?
What do you want to do?
What will happen if ... ?
What has the system done now?
What is the system trying to tell you with this message?
Why has the system done that?
What were you expecting to happened then?
What are you doing now?

Audio cassette recorder

The part of the session when participants used the program was audio recorded to provide a back up to the form for data specified above.

Evaluation questionnaire

The participants completed this questionnaire after they had completed all of the set tasks. This questionnaire was based on useful debriefing questions that have been outlined (Monk et al, 1993) and possible post session interview questions that have been suggested (Hix & Hartson, 1993). (Appendix D).

7.2.4 Procedure

For the formative evaluation session, participants were seen individually. At the start of the session, the participants were provided with the *Instructions for program evaluation.* Participants were then observed while they completed the tasks that had been set. By using the *Form for data,* notes were taken concerning critical incidents, participants' comments and relevant actions at the human-computer interface. An audio recording of the session was taken to provide a back up to this record. When the participants had completed the preplanned tasks, they completed the *Evaluation questionnaire.* Participants were de-briefed at the end of the session.

7.3 Findings

7.3.1 The tasks

Select the study 'TV violence'

All of the six students completed this task successfully by selecting the button 'TV violence'. The students were all therefore provided with the TV violence introductory screen where they could select the button presented at the bottom of the screen 'Study details' to find out more about the study.

Find out more about the study 'TV violence'

None of the six students completed this task successfully. In this first prototype, the TV violence introductory screen presented text which read "To find out more about the study

select 'Study details'", but none of the students clicked this button. For example, one student read out aloud the relevant text "select study details", but she still did not select the button 'Study details' (B2).

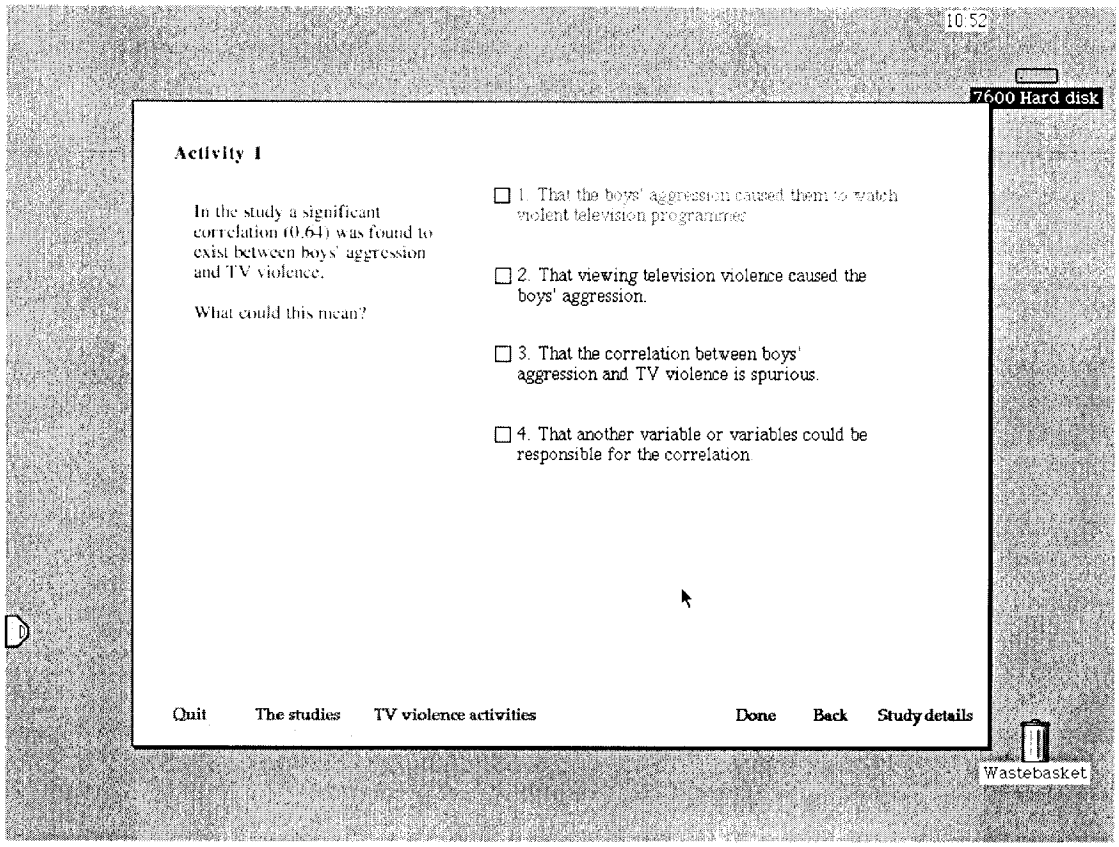
Do 'activity 1'

All of the six students successfully invoked the screen that provides the first activity, by selecting the button 'Activity 1' that is provided on the TV violence introductory screen.

For this activity, the students should interpret the meaning of a single correlation coefficient obtained in the TV violence study, by selecting possible interpretations that are provided in the form of options that can be clicked by the student. Students are then supposed to select the button 'Done' that is provided at the bottom of the screen so that they can receive audio feedback (figure 7.1). Although all of the options are possible interpretations of a correlation, all of the students initially only selected one possible option. More specifically:

- None of the six students selected the first option 'that the boys' aggression caused them to watch violent television programmes'.
- Two students selected the second option 'that viewing television violence caused the boys' aggression' (B1, B2).
- Two students selected the third option 'that the correlation between boys' aggression and TV violence is spurious' (B3, B6). One of these students did, however, select additional options after thinking that her answer may have been incorrect (B3). (See below).
- Two students selected the fourth option 'that another variable or variables could be responsible for the correlation' (B4, B5).

Figure 7.1 Activity 1



Having selected an option, three of the students selected the button 'Done' and therefore were the only ones to receive the audio feedback (B2, B3, B6). When an option had been selected, two of the students selected the button 'Back' that is provided throughout the program so that students may get back to where they have just been or review a screen prior to an action at the interface (B1, B4). These two students did not therefore receive the necessary feedback to the activity, but went back to the introductory screen for the TV violence study. When one student had decided on option 3 'that the correlation between boys' aggression and TV violence is spurious' and had clicked to select this option, she was not at all sure what to do and commented that "maybe [her] answer was wrong" (B3). She proceeded to select options 2 and 4 and then selected 'Done' and mentioned that "the thing is I wasn't sure whether the answer was correct or not" (B3).

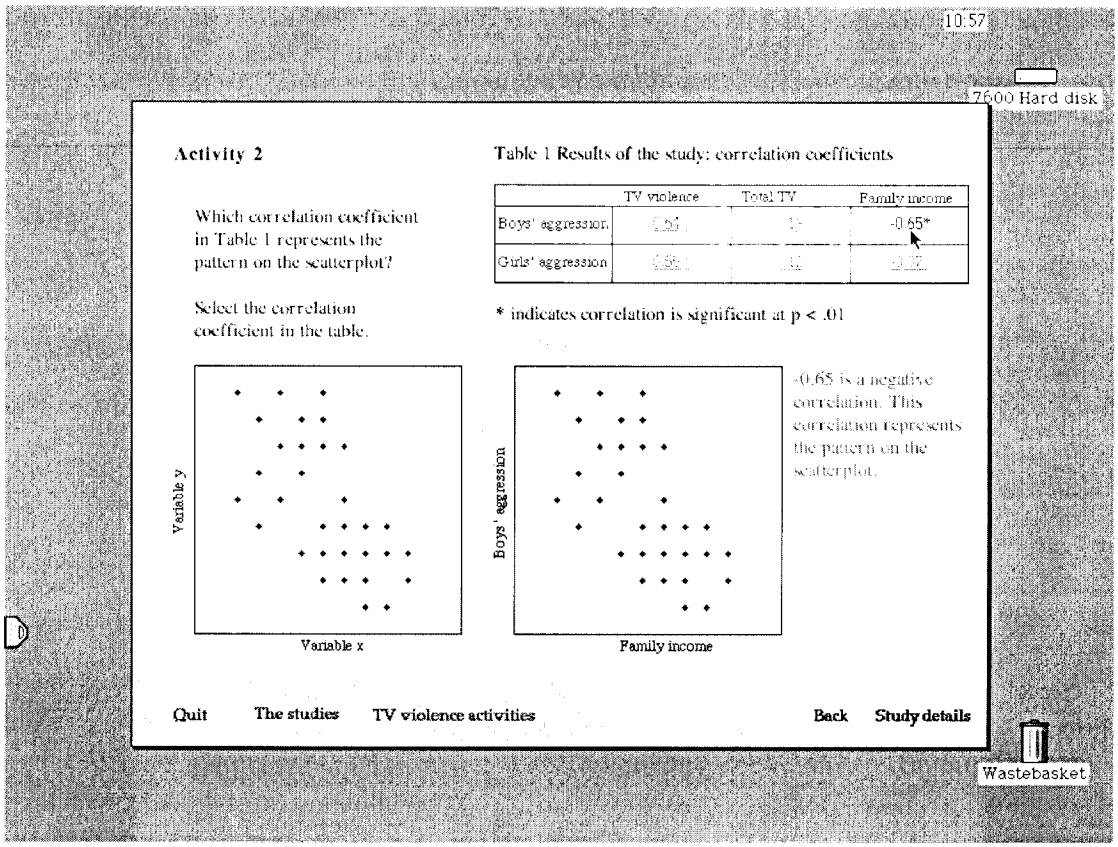
Three of the students were not clear how they could move on to, for instance, the next activity (B2, B3, B5). For example, one student having selected an option on

activity 1, asked "where shall I go?", but then selected the button 'TV violence activities' which took her back to the TV violence introductory screen that is required so that a student can select an additional activity (B5). Another student, having listened to the audio feedback to activity 1, asked "how can you move on?" before she selected the button 'TV violence activities' (B3).

Do 'activity 2'

None of the six students encountered any problems while carrying out this task. They selected correlations in the table of data and viewed the scatter plot(s) that were displayed (figure 7.2). For example, one student referred to the target scatter plot and commented "it could be [a] minus correlation" (B4) and then went on to select -0.65 in the table. When the scatter plot that represented the correlation -0.65 was displayed alongside the target scatter plot and the feedback "-0.65 is a negative correlation. This correlation represents the pattern on the scatter plot" was provided, she said "yeah I'm correct" (B4) (figure 7.2). Another student initially selected the coefficient -0.07 in the table and then went on to select the correlations 0.12 and 0.18 in turn before selecting the correct correlation of -0.65 (B2).

Figure 7.2 Activity 2. The correlation -0.65 has been selected



The study was concerned with the usability of the program, but from an instructional perspective it is noteworthy that one of the students held a particular misconception about a positive and negative correlation. With reference to scatter plots and activity 2, this student said: "if the plots gather around the line it's gonna be positive. If it's spread around it's gonna be negative" (B3). This smacks of a unidirectional conception of correlation.

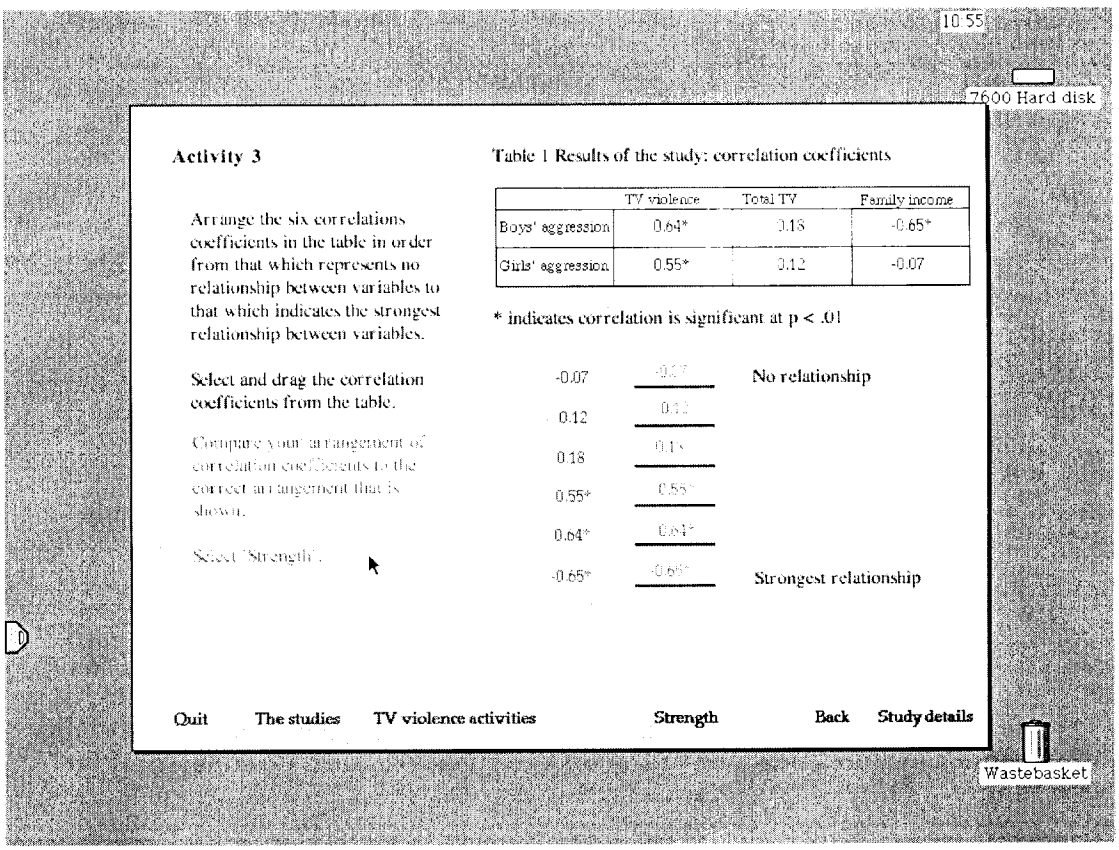
Having completed this second activity, students did not encounter any particular difficulties with moving on to complete activity three. For example, one of the students selected the button 'TV violence activities' and was therefore presented with the TV violence introductory screen from where she selected the button 'Activity 3'.

Do 'activity 3'

In this activity, students are asked to select the correlation coefficients in the table and

then drag and position them on the screen in the appropriate order from that which represents the strongest relationship to that which represents no relationship between two variables (figure 7.3). It is only necessary to click the coefficients in the table once to select them, but two of the six students double clicked the coefficients before they dragged them across the screen (B1, B6). At first, one of the students was not sure how to select the coefficients in the table and she was shown how to do this (B3). Two of the students had no trouble in selecting and dragging the coefficients from the table (B2, B5) and, when one student was dragging the coefficients from the table she remarked that it was "clever" (B6).

Figure 7.3 Activity 3



On this activity, five of the six students positioned the correlation coefficients in the correct order (B1, B2, B4, B5, B6). However, having positioned the correlations correctly, one of these students referred to the strongest coefficient -0.65 and remarked "it's a negative, so it's not the strongest relationship, but in a negative way" (B6).

Another one of these students originally positioned the correlation coefficients as follows:

-0.65 No relationship
-0.07
0.12
0.18
0.55
0.64 Strongest relationship

She then, however, referred to the correlation of -0.65 and commented "but this is not no relationship" and said "so it is probably wrong" (B1). She then re-ordered the coefficients in the correct order (B1).

The one student who positioned the correlations incorrectly on this activity, selected and dragged only two correlation coefficients to represent the strongest relationship and no relationship between variables as follows:

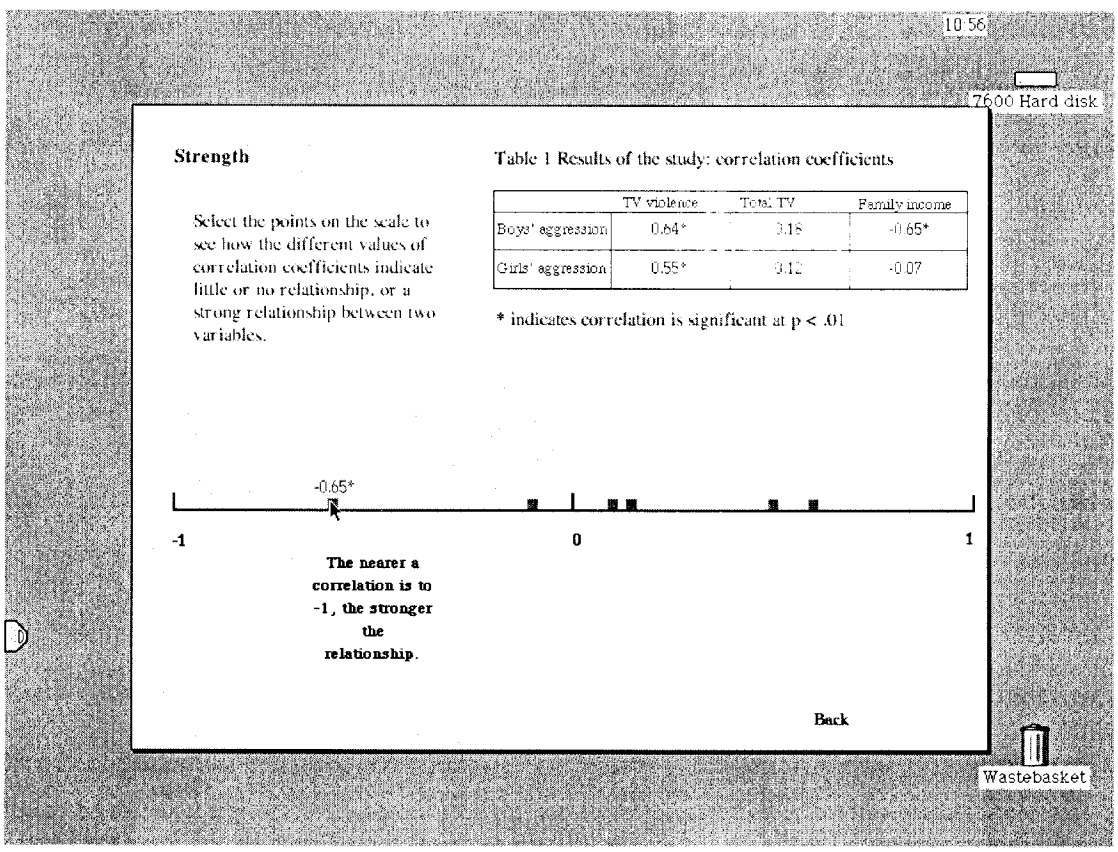
-0.07 No relationship
—
—
—
—
0.64 Strongest relationship

It was clear that she therefore thought that 0.64 was the strongest relationship amongst the coefficients in the table (B3). When this student had selected the button 'Done' and had therefore received the feedback to this activity, she commented that she "forgot about positive and negative" (B3). This student had not realised that she was meant to have positioned all of the correlation coefficients in a sequence.

When students are provided with the feedback to activity 2, text on the screen tells them to select 'Strength' (figure 7.3). If a student clicks the button 'Strength', an additional screen is invoked which provides a scale that ranges from -1 to 0 to 1. This is designed so that students can see how correlation coefficients can range from -1 to 1, and therefore indicate a strong relationship or little or no relationship between two variables.

Only half of the students selected the button 'Strength' (B2, B3, B6). One of these students simply selected the button 'Back' once she was provided with the screen that presents the scale (B2), and one of these students was not clear about what she should do with the representation and information provided by the screen in question. For example, having read the text presented on this screen she commented "I don't know what to do" (B3), and pointed to the scale on the screen and said "I've never seen this. Maybe I missed a class" (B3). Another student had difficulty selecting the points on the scale. For example, she clicked on a red point and rather than information being provided concerning the strength of a particular coefficient (figure 7.4), nothing happened in response to her clicks of the mouse (B6). This was because she had not positioned the computer mouse exactly on the red points provided on the scale.

Figure 7.4 Addition screen for activity 3



7.3.2 Evaluation questionnaire

The evaluation questionnaire that students completed when they had finished working through the program, included the following questions:

What do you think was the best thing about the program?
What do you think was the worst thing about the program?
What do you think needs changing in the program?
What did you think of activity 1 in the program?
What did you think of activity 2 in the program?
What did you think of activity 3 in the program?

When students completed this questionnaire they referred to the program that had already been set up, and looked, for example, at particular screens, such as those that provided the particular activities. In the main, the responses on the evaluation questionnaire confirmed the observations that were taken while students carried out the set tasks. For example, as outlined above, a number of students only selected one of the possible options on activity 1 and in response to "what did you think of activity 1 in the program?" one of the students wrote "I wondered whether I can [could] pick one of these or more" (B2).

Students commented on the buttons that were presented at the bottom of the screen in the first prototype of *Link*. For example, in her answer to the question "what do you think was the worst thing about the program?" one student wrote "I did not understand [what] the button 'the studies' is for" (B1), and she also commented on the questionnaire that she "didn't know which button to choose to find out more about the [study] details, so I just chose the button 'the studies'" (B1). (The button 'The studies' takes students to the introductory screen of the program). In response to the question "what do you think needs changing in the program?" one student wrote "buttons at the bottom of the screen" (B5).

Two students did not have a very high opinion of some aspects of the screen presentation, for example, of the way in which the program's instructional questions or

activities were presented. When the questionnaire asked students what they thought was the worst thing about the program, one student wrote "presentations on the screen of the questions" and this student commented that this aspect of the program needs changing (B5). In this respect, another student commented that "the font of letters, text box[es], and table look very squared" (B6). For this student, there was not an appealing presentation of text and data at the interface.

From an instructional view point, it is interesting to note that in response to "what did you think of activity 2 in the program?" where the student can see a correlation coefficient linked to the pattern it shows on a scatter plot, one student wrote that this activity was "interesting for someone learning statistics so he or she can picture the relationship between the scatter plot and the correlation coefficient" (B5), and another student's comments read "very useful to grasp the idea of how, visually, coefficients should be" (B6).

7.3.3 Additional observations

The *Instructions for program evaluation* did not specify that the participating student should not inform the other students on their course, who might also take part in the study, about the program or, more importantly, about the answers to the learner activities. Instructions were therefore modified in this respect for the second phase of the formative study that is described in chapter 8.

When this first phase of the study was carried out it became evident that students could become confused by the term task versus the term activity. Students were told that they were to complete tasks so that the program could be evaluated (e.g., the *Task sheet* specified task 1 as "select the study 'TV violence'"), but they also worked on activities when they tried out the program. The evaluation questionnaire, however, asked students "how easy did you find the tasks?" and this could be interpreted as "how easy did you find the activities?" For example, when completing the evaluation questionnaire, when the

first student to take part in the study answered the question "what did you think of activity 1 in the program?" she referred to task 1 on the *Task sheet* (B1). The difference between a task and an activity was clarified for subsequent students who took part in the first phase of the study. The *Task sheet* was also modified for the second phase of the formative study and did not therefore list the tasks as task 1, task 2, etc.

7.4 Modifications to the first prototype

The findings of the first phase of the formative study highlighted particular aspects of the program's human-computer interface that needed to be changed. These modifications were designed to improve the first prototype in terms of its usability. A second prototype of the program *Link* was therefore produced that was evaluated in phase two of the formative study.

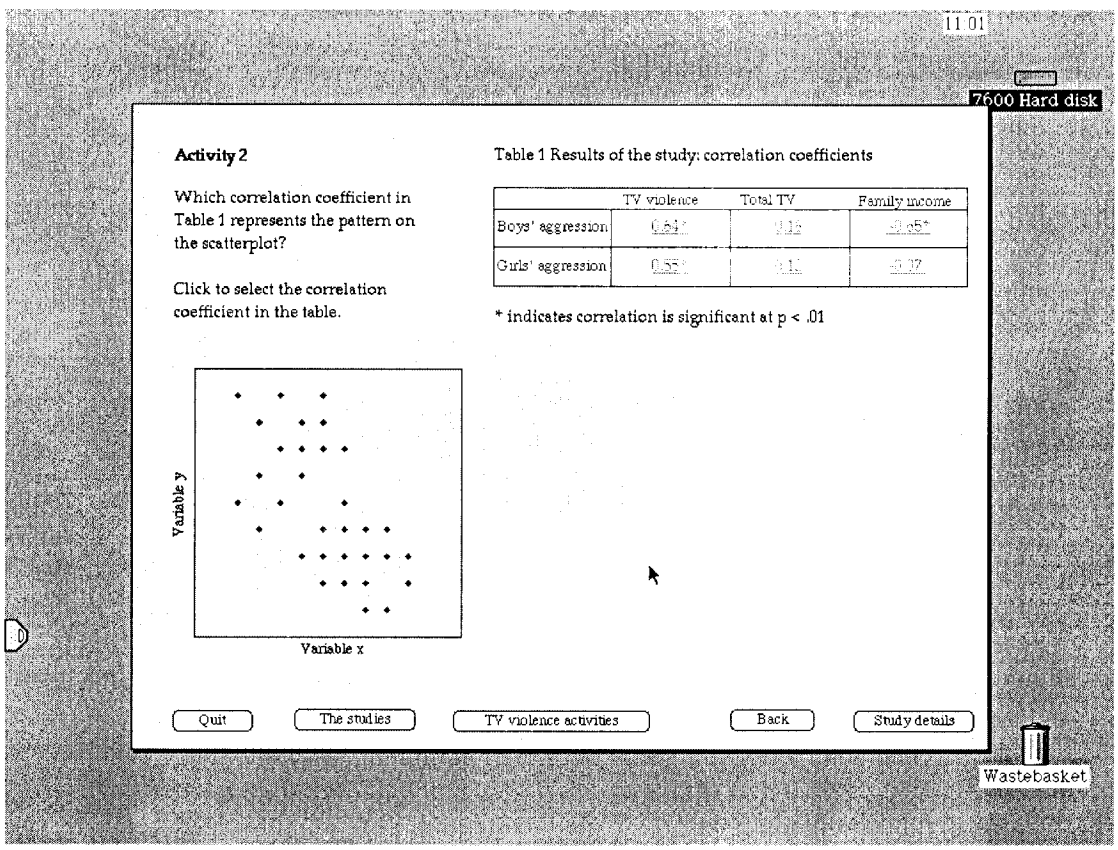
7.4.1 General modifications to the prototype

With regard to the findings outlined, the following changes were made to the first prototype:

- The buttons that were displayed at the bottom of the screen (e.g., 'The studies', 'TV violence activities', 'Study details') were changed so that they were the same style as the other buttons used in the program (e.g., the buttons 'TV violence', 'Activity 1'). This meant that they were clearer because they appeared as a button that is designed to be selected with the mouse (figure 7.2, first prototype and figure 7.5, second prototype).
- In the first program, text was presented that suggests that students should select certain buttons. For example, in activity 2 the following text, which was provided as part of the feedback to the activity, read "select 'Strength'". This aspect of the program was changed so that if a button, such as the 'Strength' button, was referred to in text then this would be explicit and would therefore be changed to read "Select the button 'Strength'".

- To improve the presentation of the activities and their layout, the grey text boxes were removed and therefore do not appear in the second prototype. In addition, the Times font (size 12) that was used in the first prototype was changed to Palatino font (size 12). This meant that the overall look of the program's interface was improved. Also, the type of font on the buttons was changed so that they were clearer to read. It was thought that removing the grey text boxes, and changing the style of font predominantly used in the program would improve the presentation of, for example, the learner activities and therefore the appearance of the program's interface. This is evident from the screen that presents activity 2 as it appears in the first prototype (figure 7.2) and as it appears in the second prototype (figure 7.5).

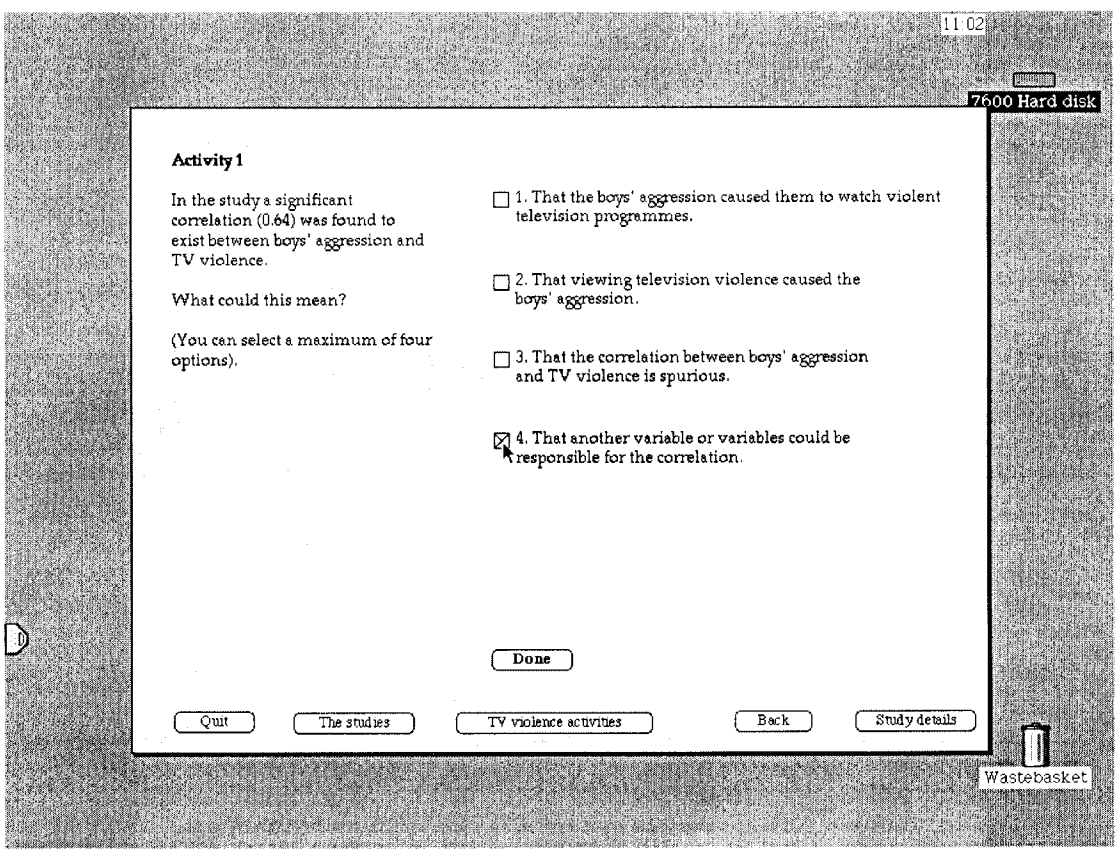
Figure 7.5 Activity 2. For the second prototype the buttons at the bottom of the screen were changed



7.4.2 Modifications to activity 1

With respect to the findings that concerned how students completed activity 1, it was decided that text on the screen should suggest to users that they can select more that one option in response to the question which asks what a correlation of 0.64 could mean. Text that reads "you can select a maximum of four options" was added to the screen that presents activity 1 (figure 7.6). To ensure that students would be provided with the necessary audio feedback to activity 1, the button 'Done' that was positioned at the bottom of the screen in the first prototype was moved so that it was displayed just under the options in activity 1. It was thought that this change would mean that students would see the button 'Done' and would therefore select it after they had selected available options in activity 1.

Figure 7.6 Activity 1 in the second prototype



7.4.3 Modifications to activity 2

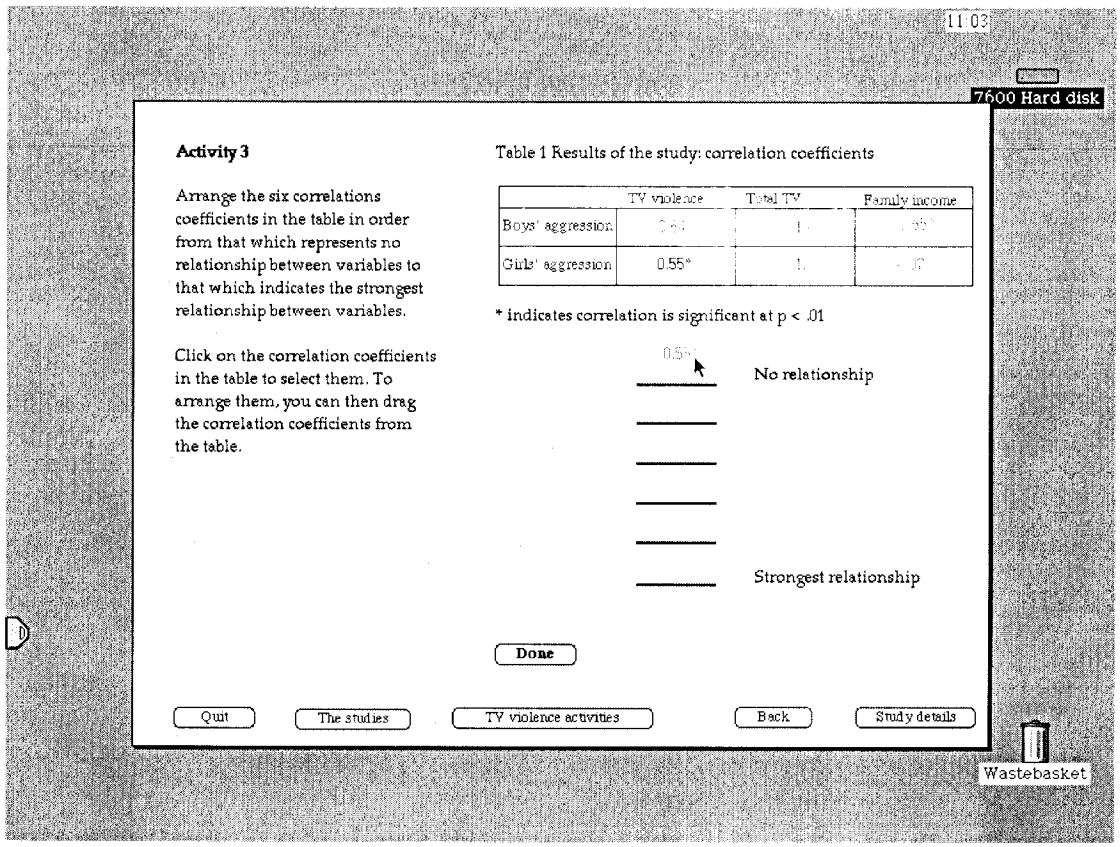
The text displayed on this activity was changed to include the following "click to select the correlation coefficient in the table". This change was necessary to ensure consistency regarding how instructions were written in the program because a similar change was needed on the screen of activity 3.

7.4.4 Modifications to activity 3

To ensure that students did not click twice on the correlation coefficients displayed in the table for activity 2, the instructions for this activity were revised for the second prototype: "Click on the correlation coefficients in the table to select them. To arrange them, you can then drag the correlation coefficients from the table".

A number of students did not select the button 'Strength' in this activity so it was modified in accordance with the changes that concerned the buttons presented at the bottom of the screen, which were outlined above. In addition, the position of this button 'Strength' was changed so that it was presented immediately under the sequence of correlation coefficients arranged by the student. If the student selects the button 'Strength', the screen that presents a scale for a correlation coefficient is provided. The text on this screen was also revised for the second prototype so that the student would have a clearer idea of what they should do on this screen: "Click on the red points on the scale. You can then see how the different values of correlation coefficients indicate little or no relationship, or a strong relationship between two variables" (figure 7.4, first prototype and figure 7.7, second prototype).

Figure 7.7 Activity 3 in the second prototype



7.5 Summary

This chapter described the first phase of the formative evaluation study that was conducted to assess the usability of the first prototype of *Link*. This study involved psychology students who tried out the program in a computer laboratory at their university. The study provided qualitative data concerning the usability of the program and in the light of such data, a revised prototype was produced to be used in the second phase of the formative evaluation study. This second phase is described in the following chapter.

Chapter 8

Formative evaluation: phase two

8.1 Introduction

This chapter describes the second phase of the formative evaluation study that was conducted. The evaluative framework that was outlined in chapter 6 was used in this empirical study and the study involved students who were studying psychology at an Open University residential school. The primary aims of this study were to:

(a) Investigate whether *Link* contributed to students' general understanding of correlation.

(b) Find out whether *Link* affected students' conceptions in correlation.

(c) Provide a formative evaluation of the program's learner activities and related presentation of topic material.

(d) Pilot tests that were designed to provide an assessment of students' understanding of correlation.

An expert evaluation of *Link* was also carried out and this chapter describes the outcome of this evaluation. The findings of the formative and expert evaluations were used to inform the development of a final version of *Link*, which is described in chapter 9.

8.2 Method

8.2.1 Design

A formative evaluation of the second prototype of *Link* was conducted with target users at a residential school.

8.2.2 Participants

Eighteen students (nine females and nine males) who were studying part time for an undergraduate psychology course with the Open University took part in the second phase of the formative study. The mean age of this group of students was 38 years (mean = 37.82, S.D. = 7.51, minimum = 24, maximum = 56). Fifteen out of the eighteen students had an O' level or GCSE (grade C or above) in Mathematics. Out of the total number of students, seventeen of them had completed an introductory psychology course that covered the topic of correlation, and one of the students had completed a degree course in psychology.

With regard to computing experience, the average number of years for having used computers was five years (median = 5, minimum = 0 years, maximum = 33 years). Nine of the participants had used both an Apple Macintosh and an IBM PC-compatible, five had only used an IBM PC-compatible, three had only used an Apple Macintosh, and although only one student had used neither of these computers, they had previously used an Amstrad 9512 for word-processing. All but one of the eighteen participants reported that they used the computer for word-processing, thirteen of these using Microsoft Word and four of these using another application, such as Word Perfect. The participants reported that they made fairly regular use of the computer: nine participants indicated that they used it every day, five participants used it every two to three days, one participant used it once a week, and three used it less than once a month.

To ensure anonymity the identifier (e.g., O1) is used for the different participants.

8.2.3 Materials

Hardware and software

The second prototype of the program was stored on a Apple PowerBook 520c from which it was run. The screen of the PowerBook was at the required 640 x 480 resolution and 256 colours. A mouse was attached to the PowerBook for participant use.

Participant profile

This questionnaire was completed by the participants to collect data concerning, for example, the participants' gender, age, qualifications, the university courses they had taken, and their computing experience. Part of this questionnaire was based on a sample Computer Experience Questionnaire (Draper et al, 1994).

Instructions for program evaluation

This outlined the purpose of the session and told the participants what would happen during the session. These instructions also told the participants that they were to be observed while they used the program. They were informed that they were to complete a set of tasks and were asked to think aloud while they did this.

Task sheet

This sheet asked the participant to complete the following tasks:

- Select the study 'TV violence'.
- Find out more about the study 'TV violence'.
- Do 'activity 1'.
- Do 'activity 2'.
- Do 'activity 3'.

Form for data

This provided a schedule to collect qualitative data, such as participants' actions and comments in relation to the specified tasks.

Audio cassette recorder

The part of the session when participants were using the program was audio recorded to provide a back up to the *form for data* specified above.

Program evaluation questionnaire

After the participants had completed all of the set tasks, they completed a modified version of the evaluation questionnaire that was used in the first phase of the formative study. This questionnaire was designed to elicit participants' subjective opinions about the program.

Tests in correlation

Two equivalent tests in correlation were developed. The questions in these tests were used in the investigation described in chapter 3, and were adapted for the formative study. Questions that were based on exercises from particular texts (Coolican 1990; Gravetter & Wallnau, 1995a; Gravetter & Wallnau, 1995b; Pagano, 1990; Shavelson, 1981; Weinberg & Goldberg, 1990), were also used in the tests.

These tests were designed to assess students' understanding of correlation and comprised of twenty five questions. Filler questions were included in the tests to ensure that participants would not determine the purpose of the tests. Each of the tests provided four quantitative measures:

- Overall score (out of a total of 25).
- Overall score that excluded the filler questions (out of a total of 19).

- Whole question score (out of a total of 14).
- Whole question score that excluded the filler questions (out of a total of 10).

The questions were also devised to identify students' misconceptions in correlation. Each of the equivalent tests included:

- Four questions designed to identify a causalistic conception of correlation.
- Seven questions designed to identify a unidirectional conception of correlation.
- Eight questions designed to identify the conception that a positive correlation is stronger than a negative correlation when this is not the case.
- Six filler questions that concern the topic of correlation.

The questions for the tests were examined for accuracy and clarity by two subject specialists and were modified accordingly for the developed tests (appendix E).

8.2.4 Procedure

For the evaluation session, participants were seen individually. At the start of the session, the participants were provided with the *Instructions for program evaluation.* During the session, participants completed two equivalent tests in correlation (test A, test B). Participants were randomly assigned to either complete test A prior to using the program and to complete test B after they had used the program, or vice versa. Participants were observed while they completed the tasks that had been set. By using the *Form for data,* notes were taken concerning, for example, participants' comments and relevant actions at the human-computer interface. An audio recording of the part of the session when participants were using the program was taken to provide a back up to this record. When the participants had completed the preplanned tasks, they completed the *Program evaluation questionnaire.* Participants then completed the other test in correlation and were de-briefed at the end of the session.

8.3 Findings

The approximate time that it took participants to complete the pre test, post test and the tasks or the time taken to work through the program was noted. The pre test and post test each took approximately between ten and twenty minutes to complete. It took participants between ten and fifteen minutes to carry out the specified tasks.

8.3.1 Interactions: the program

The findings that concern the participants working with the program are described with respect to the five tasks that the participants completed. In addition, a number of issues that concerned the usability of the program became apparent while observing participants working through the program. These generic issues will be considered separately.

The first task asked participants to select the study 'TV violence'. All of the eighteen participants successfully completed this task by selecting the button 'TV violence'.

Find out more about the study 'TV violence'

In completing this task, participants were expected to select the button 'Study details' because the text on the introductory screen for the study TV violence reads "to find out more about the study select 'Study details'." However, four of the participants did not select this button at this point in the program (O3, O5, O8, O17). One of these students did go on to select the button 'Study details' when they were completing activity 1.

Both the introductory screen to the TV violence study and the screen that presents details of the study provides data in the form of a table. Six of the fourteen participants who did find out more about the study by selecting the button 'Study details', had difficulty interpreting the table (O3, O4, O9, O10, O12, O14). For example, one of the participants referred to the table and said "I'm confused as to just what this is telling me" (O12). (See also case studies two and three).

Do 'activity 1'

As previously described, with this activity students are meant to interpret the meaning of a correlation coefficient by selecting four possible options. When participants completed this activity, a variety of interactions were observed:

- One participant selected the first option only, 'that the boys' aggression caused them to watch violent television programmes' (O18).
- One participant selected the second option only, 'that viewing television violence caused the boys' aggression' (O1).
- One participant selected the third option only, 'that the correlation between boys' aggression and TV violence is spurious' (O14).
- Three participants selected the fourth option only, 'that another variable or variables could be responsible for the correlation' (O4, O5, O8).
- One participant selected the second and fourth options (O2).
- One participant selected the third and fourth options (O7). This student provided an explanation as to why she did not select the first and second options. She initially referred to the first option and commented "number one. It didn't show that at all because the word cause is in it and just 'cos there's a correlation doesn't mean there's a cause and effect. The same for number two" and that "there's nothing to say that the watching the programme caused the aggression. It could be something

else that caused the aggression it just happen[s] that the two things correlate" (O7).

- Two participants selected the first, second and fourth options (O6, O13).
- Four participants selected all of the four options (O9, O10, O11, O15).

Four participants did not select any of the options, but these users did select the button 'Done' and therefore received feedback to the activity (O3, O12, O13, O17). This did not necessarily mean that they did not, for example, read through the options: one of these participants referred to each of the options in turn. To the third option she said "not that one" and then having read the fourth option decided "no, I think it means it's either one or two" (O17). One of these participants referred to the options and simply said "it could be all of them" (O3).

Two of the participants commented that they did not know the meaning of the word spurious that is included in the third option 'that the correlation between boys' aggression and TV violence is spurious'.

All but one of the participants selected the button 'Done' and were therefore provided with the audio feedback to the activity. This feedback asks students to think what kind of variable could be responsible for the obtained correlation between the boys' aggression and the TV violence. Three of the participants were unsure about what to do in response to this feedback. For example, one participant asked "am I supposed to be answering this now?" (O15), and another commented "I'm not sure what it wants me to do here" (O13). In spite of this, thirteen out of the eighteen participants provided an answer to the feedback. The participant comments below illustrate the kinds of responses that were provided to the audio feedback of this activity: "what kind of third variable could be responsible for the relationship that was found to exist between boys' aggression and TV violence?"

"Well, there's all sorts of variables ... in terms of parental example" (O11).

"Is she repeating back what I said or is she explaining it to me? ... Does she want

me to answer her? 'Cos I can do that if you like lady... Home background, locality, where they live, whether they like football, god knows, any number of things. Maybe they ate too much sugar and they've got hyperactive. All sorts of things" (O10).

"Maybe they are in a more aggressive home and that they've got older siblings who are aggressive and therefore they are seeing it and that makes them more aggressive by imitation perhaps" (O9).

"Could be parental attitude toward child rearing" (O5).

Do 'activity 2'

For activity 2 participants are expected to select the coefficient in the table (-0.65) that represents the target scatter plot that is displayed on the screen. Only two of the participants selected this coefficient, -0.65 almost immediately (O7, O9). Apart from these two cases, a variety of approaches to activity 2 were observed.

At the beginning of the activity, four out of the eighteen participants appeared not to be able to attempt the activity, but they then completed it and worked out that the negative correlation of -0.65 represented the scatter plot in question (O3, O6, O11, O15). For example, one of these participants said "I'm puzzled, I'm afraid" and was prompted to select a correlation as instructed by the program. She selected the correlation 0.64, and then commented "Oh no, so it's the other way" and selected the correlation -0.65 (O15). Another participant approached the activity in a similar way (O11). Initially, this participant remarked "don't like these scatter graphs" and went on to say that she could not "picture it in a graph form like that" (O11). This participant then said she would guess and selected the correlation 0.64 and said "so that's [a] positive correlation" and "so we need a negative correlation" and she then selected -0.65. This participant did, however, comment further "I have to say I still don't understand why it represents what it says it does on the graph" (O11). In completing activity 2, one participant commented "I haven't the first idea ... I'm completely lost here ... I could have a guess", but then selected the

positive correlation 0.55 in the table. A scatter plot that represents this coefficient was displayed on the screen and this participant said "in that case it must be that I'm looking for a negative correlation" and selected -0.65. He then, however, added "that's not a bad guess second time ... I'm not familiar with scatter plots" (O6).

Four of the participants evidently found activity 2 difficult and tended to guess by selecting several of the correlations in the table until they selected -0.65, which is the one that represented the target scatter plot (O5, O10, O14, O18). For example, one of these participants said "I haven't a clue. I've never clapped eyes on a scatter plot before" and remarked how he was guessing when he selected the following correlations in the table in turn: -0.07, 0.12, 0.64, -0.65 (O5). Similarly, one participant decided "so I'll go for the nearest one at random" and selected the correlations 0.55, 0.64, 0.18 and then -0.65 in turn (O18).

Four of the participants did not mention that they found the activity difficult and that, for example, they were stuck or confused, but they did select several correlations in the table including the negative correlation that represented the target scatter plot (O8, O12, O13, O16). For example, one of these participants referred to the target scatter plot and said "it's not scaled it could be anything" and then selected the correlation 0.64, 0.18 and -0.65 in turn (O13). Another one of these participants selected the correlations 0.18, 0.12, -0.07, 0.55 in turn and then said "tried four out of six now. It's none of them" and then selected the coefficient of -0.65 (O16).

Do 'activity 3'

Thirteen out of the eighteen participants successfully completed this activity by selecting the correlation coefficients with the mouse and dragging them from the table and arranging them on the screen in the correct order as follows:

-0.07 No relationship
0.12
0.18
0.55
0.64
-0.65 Strongest relationship
(O1, O2, O6, O7, O8, O9, O10, O11, O12, O14, O15, O17, O18).

When completing this activity, one of these participants remarked "I can't remember if I've got this right or not" (O15).

Five of the participants did not arrange the correlation coefficients from that which represents no relationship to that which represents the strongest relationship. Indeed, three out of these five participants positioned the positive correlation 0.64 as representing the strongest relationship (O3, O5, O16). (One of the five participants found the screen of this activity very confusing and had difficulty carrying out the activity, O13). More specifically:

- Three participants arranged the correlation coefficients as follows:

  -0.07 No relationship
  -0.65
  0.12
  0.18
  0.55
  0.64 Strongest relationship
  (O3, O5, O16).

- One participant arranged the coefficients as follows:

  0.64 No relationship
  0.55
  0.18
  0.12
  -0.07
  -0.65 Strongest relationship
  (O4).

Two of the participants did not select the button 'Done' when they had finished arranging the correlation coefficients which meant that feedback in the form of text was not provided on the screen (O1, O8). The feedback that is provided to this activity informs the user to select the button 'Strength'. Four out of the eighteen participants did not select this button and so did not invoke the screen that provides a scale designed to illustrate that a correlation can vary from -1 to 0 to 1. When one of these four participants had attempted to complete activity 3, it was suggested that they finish using the program because due to health problems it was not appropriate for them to continue.

Fourteen out of the eighteen participants studied this last screen that provided a scale to illustrate how a correlation coefficient can indicate a strong or weak relationship between two variables. These participants selected the red points on the scale, which invoke different values of correlation coefficients (e.g., -0.07), and read the text that was subsequently displayed (e.g., "a correlation that is near 0 indicates little or no relationship"). With regard to this screen, one participant pointed out "I'm reading the instructions about strength and ... they don't seem very clear to me". (O14). One participant, who had correctly arranged the correlations in activity 3, tried out the scale and remarked that he must have put -0.65 in the wrong place on the activity. It is important to point out that there was an error in one section of text that was displayed when the user selects a particular point on the scale: the text read "the nearer a correlation is to 1, the stronger the relationship" when it should have read "the nearer a correlation is to a -1, the stronger the relationship".

The usability of the program

It was outlined above that on activity 1 one participant (O1) did not select the button 'Done' and on activity 3, two participants (O1, O8) did not select the button 'Done'. This meant that these participants did not receive feedback to these activities. Nine out of the eighteen participants experienced additional problems with the button 'Done' while they used the program (O4, O6, O7, O8, O9, O11, O13, O14, O17). For example, on activity 1 when the user selects the button 'Done' audio feedback is provided, but when this feedback has played the button remains on the screen. It is faded to indicate that it is not active. However, nine of the participants selected 'Done' again after they had listened to the feedback and found that nothing happened when they did this. Two of these participants then went on to select the button 'Quit' that is meant to be used when a student wants to quit the application (O13, O14).

It was not entirely clear to all of the participants what they should do when they had completed an activity or when and how they should proceed to, say, the next activity. For example, having completed activity 1, one participant said "I don't know what to do at this point" (O10), and another participant commented "I don't know if that's the end of activity 1 or if there is more to come" (O18). This participant also experienced problems in navigating through the program when they were completing the second task that asked students to find out more about the study.

8.3.2 Outcomes: tests in correlation

Both the pre and post tests that were designed to assess a student's understanding of correlation provided an overall score for each participant. This overall score was obtained by scoring every question on the tests as correct or incorrect. For example, if a participant's answer to question 2 was correct they would receive a score of 1, but if their score to question 3(i) was incorrect that would receive a score of 0. For the pre test and the post test, the participant's scores for each of the questions were summed to obtain an

overall score. The descriptive statistics for the pre and post tests are provided in table 8.1. The maximum possible overall score for each test was 25.

Table 8.1 Descriptive statistics for overall scores of pre and post tests

	Pre test (n = 18)	Post test (n = 18)
Mean	16.22	18.33
S.D.	5.50	4.28
Minimum	7.00	7.00
Maximum	25.00	25.00

Table 8.2 Descriptive statistics for overall scores of pre and post tests (filler questions excluded)

	Pre test (n = 18)	Post test (n = 18)
Mean	11.39	13.56
S.D.	4.71	3.94
Minimum	3.00	4.00
Maximum	19.00	19.00

The scores for the filler questions that were included in both tests were excluded to provide the descriptive statistics for the pre and post tests that are summarised in table 8.2. Excluding the participants' scores on the filler questions, the maximum possible overall score for each test was 19. It can be seen from the histograms in figures 8.1 and 8.2 that there is a difference in the distribution of participants' overall scores for the pre and the post test.

Figure 8.1 Histogram of participants' pre test scores (excluding filler questions)

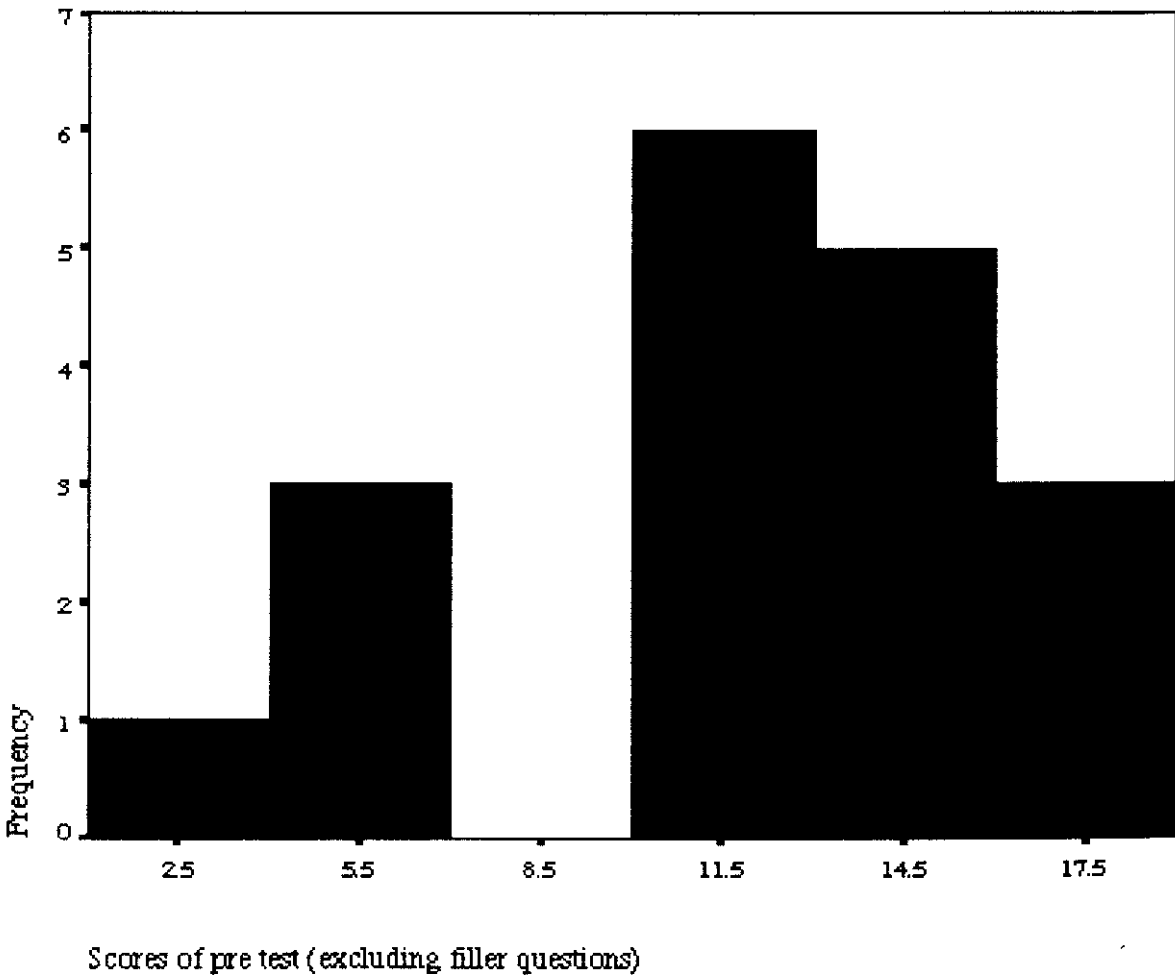
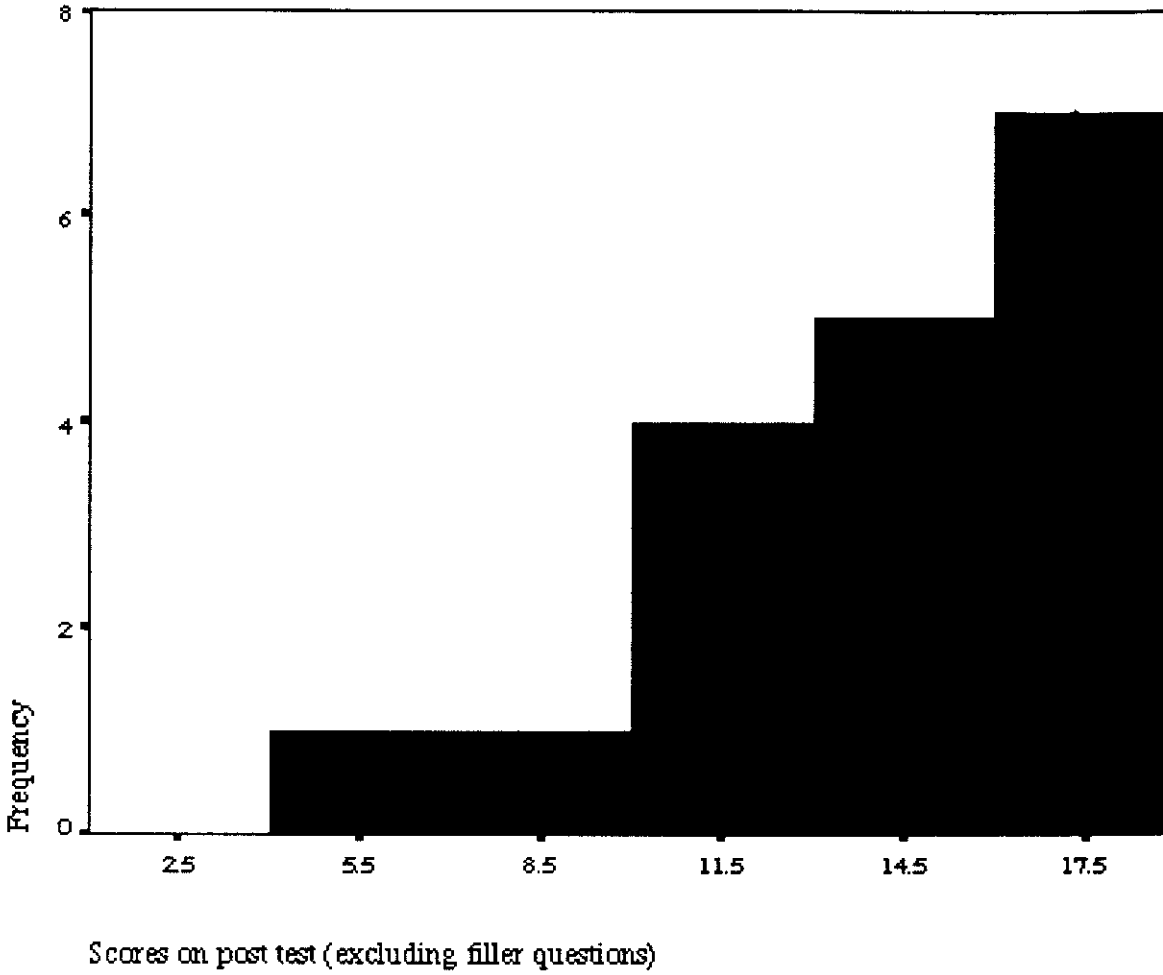


Figure 8.2 Histogram of participants' post test scores (excluding filler questions)



A one-tailed related t-test was used to see if there was a difference in the mean scores for the pre and post test. There was a significant difference between the mean scores of the pre and post test ($t = 2.22$, d.f. = 17, $p < 0.05$). There was also a significant difference between the mean scores of the pre and post test if the filler questions were excluded from the participants' overall scores ($t = 2.59$, d.f. = 17, $p < 0.05$). This indicates that the participants' post test scores were higher than their pre test scores.

To ascertain whether the tests in correlation were equivalent two-tailed independent t-tests were carried out to see if there was a difference between the mean scores of test A

and test B on the pre test. There was no significant difference in the pre test mean scores of test A and test B ($t = 1.12$, d.f. = 16, $p > 0.05$). Excluding the filler questions from the pre test scores, there was no significant difference in the mean scores of test A and test B ($t = 1.39$, d.f. = 16, $p > 0.05$).

In the equivalent tests in correlation eleven of the questions asked the participant to explain or justify their answer to a previous question. For example, question 8 asked:

8. Which correlation is stronger?
(a) -0.88
(b) 0.02
and question 8(i) asked the participant to
8. (i) Explain your answer.

In the above analysis of the pre and post test scores, participants could, for example, obtain a score of 1 (correct) or 0 (incorrect) for question 8 and a score of 1 or 0 for question 8(i). This method of scoring provided the participants' overall scores on the tests. The tests in correlation were also coded and scored for further analysis in a different way because it was clear that in some cases a participant could be correct on, for example, question 8, but provide an incorrect or insufficient explanation to their answer on question 8(i). Therefore questions of the above format that were comprised of two parts were scored as correct or incorrect as a whole question and therefore gave a score of 1 or 0. This meant that participants' answers were scored as correct only if they both answered the first part of the question correctly and provided an appropriate explanation for their answer.

Additional scores (whole question scores) of the participants were therefore calculated for the pre test and post test. The maximum possible whole question score was 14. Descriptive statistics of this measure are provided in table 8.3.

Table 8.3 Descriptive statistics for whole question scores of pre and post tests

	Pre test (n = 18)	Post test (n = 18)
Mean	8.22	9.67
S.D.	3.34	3.09
Minimum	2.00	2.00
Maximum	14.00	14.00

For the participant's whole question scores, the scores for the filler questions were excluded to provide statistics that are summarised in table 8.4. If the filler questions were not included in the whole question scores, the maximum possible score was 10.

Table 8.4 Descriptive statistics for whole question scores of pre and post tests (filler questions excluded)

	Pre test (n = 18)	Post test (n = 18)
Mean	5.00	6.39
S.D.	2.81	2.73
Minimum	0	0
Maximum	10.00	10.00

Figures 8.3 and 8.4 illustrate the difference in the distribution of participants' whole question scores in the pre and post tests (excluding the filler questions).

Figure 8.3 Histogram of participants' pre test whole question scores (excluding filler questions)

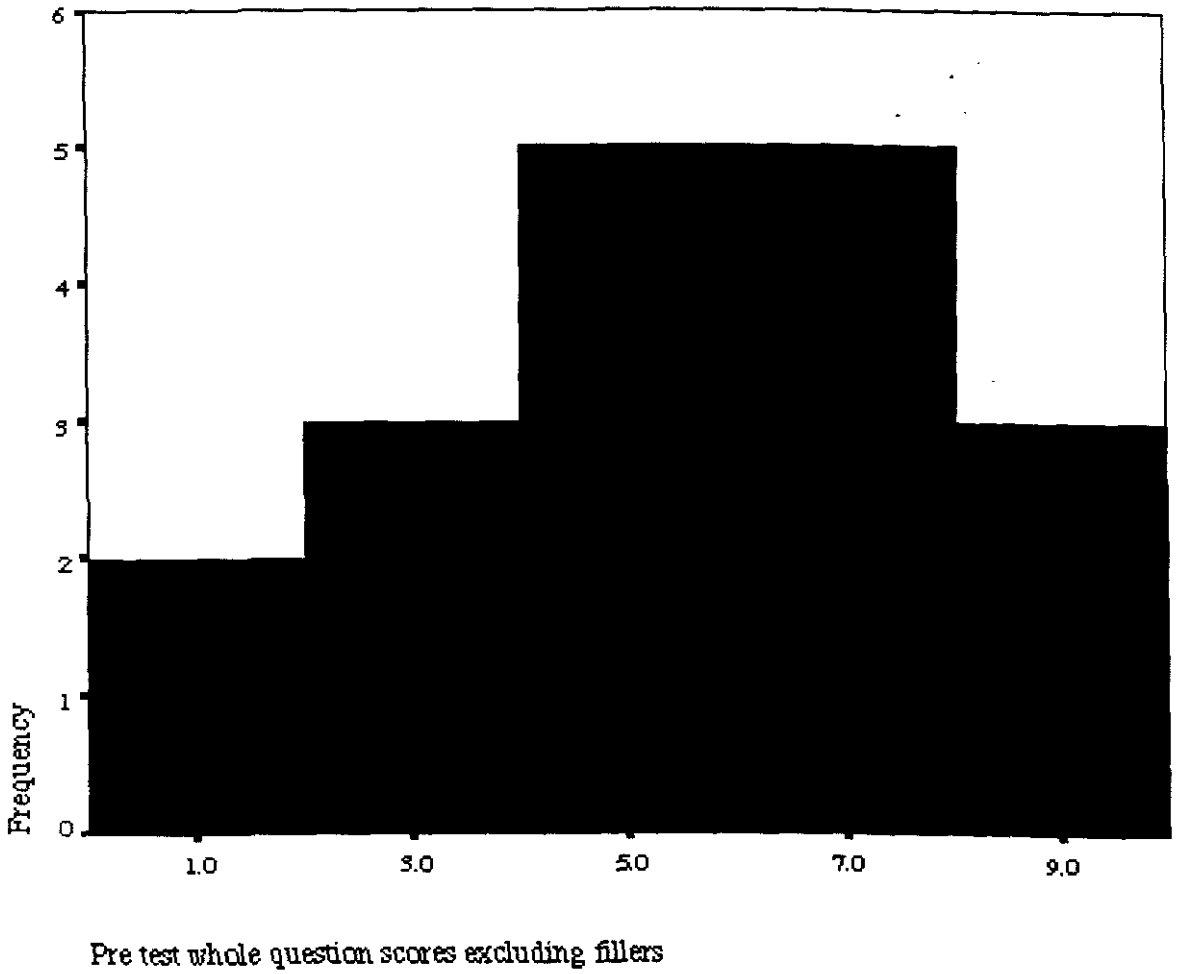
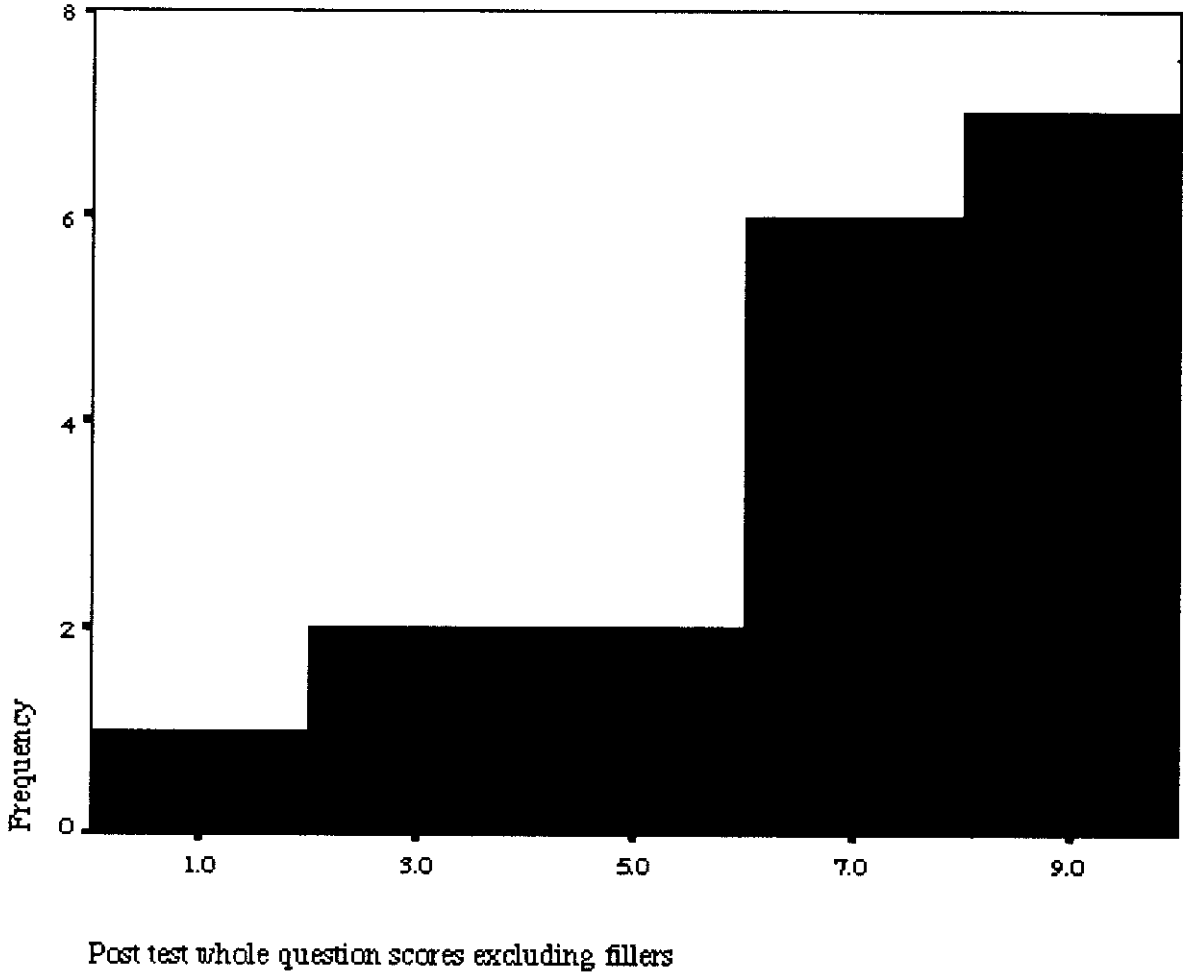


Figure 8.4 Histogram of participants' post test whole question scores (excluding filler questions)



A one-tailed related t-test was performed to see if there was a significant difference between the means of the whole question scores of the pre test and the post test. It was found that there was a significant difference in these means ($t = 2.10$, d.f. = 17, $p < 0.05$). If the filler questions were excluded from the whole question scores, there was a significant difference between the means of the pre test and the post test ($t = 2.31$, d.f. = 17, $p < 0.05$). This indicates that the participants' post test scores were higher than their pre test scores.

One-tailed Wilcoxon signed-ranked tests were also carried out to see if there was a significant difference between the pre and post test means of the whole questions scores.

Wilcoxon tests were carried out because in the use of a t-test it is assumed that the sampling distribution is normally distributed and it was thought that this was not clearly the case with the distribution of participants' whole question scores (e.g., figure 8.4). However, the results of the Wilcoxon tests showed that there was a significant difference between the pre and post test mean whole questions scores ($T = 28.5$, $n = 15$, $p < 0.05$) and a significant difference between these mean scores if the filler questions were excluded ($T = 21$, $n = 14$, $p < 0.05$).

To find out whether the tests in correlation were equivalent two-tailed independent t-tests were carried out to see if there was a difference between the means of the whole question scores of test A and test B. With respect to the pre test whole question scores, there was no significant difference between the means of test A and test B ($t = 0.99$, d.f. $= 16$, $p > 0.05$). If the filler questions were excluded from the whole question scores, no significant difference was found between the means of test A and test B on the pre test ($t = 1.19$, d.f. $= 16$, $p > 0.05$).

8.3.3 Outcomes: questions in correlation

The study was designed to see if the program contributes to a change in students' misconceptions in correlation. Participants' answers to particular questions were examined to see if there was any evidence for change. For example, question 3 on the tests was designed to tap a student's unidirectional conception of correlation and their responses to this question on the pre test and post test were examined. It could then be determined if a student held a unidirectional conception of correlation, and if so, whether this conception was evident or not in their response to question 3 on the post test. The filler questions were not therefore included in this analysis.

With the exception of question 3 and disregarding the filler questions, questions on the tests consisted of two parts. In these cases, the questions were treated as one question and the participant's answers were categorised accordingly. For example, a participant's

answers to both question 2 and question 2(i) were examined together to see if they held a causalistic conception in correlation. For simplicity, these kinds of questions, such as question 2 and question 2(i) will be referred to as question 2.

For particular questions, one-tailed McNemar tests were carried out to see if there were any significant pre and post test changes in the observed frequencies of those responses that could be categorised as a particular misconception. When the total number of changes is less that 10, the binomial test is used (Siegel & Castellan, 1988). An alpha level of 0.05 was used for all the McNemar tests.

Causalistic conception of correlation

Question 2 and question 6 on the equivalent tests in correlation were designed to identify a student's causalistic conception of correlation (appendix E). Here, the observed frequencies of a causalistic response might change from the pre test to the post test.

On the pre test, eleven of the participants' responses to question 2 were categorised as normal (O1, O3, O4, O6, O7, O9, O10, O11, O14, O17, O18) and four of the participants' responses were categorised as causalistic (O8, O12, O13, O16). However, three of the participants' responses could not be categorised as normal or as causalistic (O2, O5, O15). For example, one participant (05) answered question 2 as follows:

> 2. Professor Smith does an experiment and establishes that a correlation exists between variables A and B. Based on this correlation, she asserts that A is the cause of B. Is this assertion correct?
> (a) No (participant indicated a).
> (b) Yes
> 2. (i) Explain.
> "No significance stated. No indication of what direction correlation is in".

In this case, it was not possible to categorise the answer as normal because the explanation to the question was not sufficient.

Twelve of the participants' responses to question 2 on the pre test and/or the post test could be categorised as either a causalistic conception or a normal conception in which causality is not inferred from an obtained correlation. The results of the McNemar test are not reported here because only two of the participants' responses changed: one of the participant's responses was categorised as causalistic on the pre test and as normal on the post test, and one of the participant's responses was categorised as normal on the pre test and as causalistic on the post test.

On question 6 on the pre test, however, only four of the participants' responses were categorised as normal (O4, O7, O9, O18), eight of the participants' responses were categorised as causalistic (O5, O6, O8, O10, O13, O15, O16, O17) and six could not be placed in either of these categories (O1, O2, O3, O11, O12, O14). Participants' responses to question 6 on the post test revealed a different pattern: six responses were categorised as normal (O1, O8, O9, O10, O11, O14), five responses were categorised as causalistic (O2, O3, O5, O7, O15) and seven responses were not categorised as normal or as causalistic because, for example, an insufficient explanation was provided. For question 6, twelve of the participants' answers could not therefore be categorised as either normal or as causalistic on either the pre test and/or the post test. The results of the McNemar test are not reported because only two of the participants' responses were categorised as causalistic on the pre test and normal on the post test (and one participant response was categorised as normal on the pre test, but as causalistic on the post test).

It is noteworthy that on test A, question 6 says that a correlation has been obtained between a measure of intelligence and a measure of creative thinking and on test B, question 6 provides a correlation that has been obtained between the length of time a person is in prison and the amount of aggression a person displays on a psychological inventory. The findings above suggest that for this kind of question a participant is more likely to infer causality from correlation or to provide a response that is not possible to categorise as normal or as causalistic, than a participant is in the case of question 2.

Unidirectional conception of correlation

Out of the four questions that were designed to identify a unidirectional conception of correlation, only question 3 and question 10 uncovered this conception. In the case of question 10, only one participant response was categorised as a unidirectional conception (O5). For question 3, which was adapted from previous research that is described in chapter 3 (Morris, 1997), eleven of the participants' answers could be categorised as normal or as a unidirectional conception of correlation on the pre and/or post test. Four of the participants' responses were categorised as a unidirectional conception of correlation either on the pre test or on the post test. However, in this case the findings of the McNemar test are not reported because only two of these participants' responses changed from being categorised as unidirectional on the pre test to normal on the post test.

The conception that a positive correlation is stronger than a negative correlation

Questions 4, 8, 11 and 14 on the tests were designed to identify the conception that a positive correlation is stronger than a negative correlation when this is not the case. This kind of conception is evident if students are asked to decide, in terms of the strength of correlations, the appropriate order of a set of correlation coefficients and they answer in either of the following ways:

(i) The student will order a set of correlation coefficients as follows:

Strongest weakest
0.83, 0.65, -0.91, 0.03

Here, the coefficient that represents no correlation (0.03) is weaker than a strong negative correlation (-0.91).

(ii) The student will order a set of correlation coefficients as follows:

Strongest weakest

0.83, 0.65, 0.03, -0.91

In this case, the negative correlation is weaker than the correlation that represents no correlation (0.03).

In both cases, the student might view a positive correlation as stronger than a negative correlation, but they might view little or no correlation as stronger than a negative one or not. The tests in correlation were designed to pick up on these possibilities. However, participant responses were categorised as the strength misconception on only two out of the four questions designed to identify this conception.

In the case of test A, question 8 asked participants to judge which of the two correlations (a) -0.82 or (b) 0.04 was stronger and to explain their answer. On the pre test, twelve out of the eighteen participants gave a correct response to this question and provided an appropriate explanation to their response. However, five of the participants' responses could not be categorised on question 8, and only one participant response to this question was categorised as the conception that no correlation is stronger than a negative one. This student indicated on question 8 that 0.04 was stronger than -0.82, and justified his answer by writing "A is a negative correlation" (O13). On the post test, only one participant indicated that 0.04 was stronger than -0.82 *and* explained that it was "closest to +1" (O18). This is noteworthy because on the pre test this participant had provided the correct response to question 8 and commented that the negative "correlation coefficient is closest to -1" (O18).

Question 14 asked participants to choose, from four sets of correlation coefficients, the set that shows the weakest to strongest relationship. In one of these sets no correlation is viewed as weaker than a negative correlation, but a positive correlation is viewed as stronger than both:

Strongest weakest
0.83, 0.65, -0.91, 0.03

In another set, the negative correlation is viewed as weaker than both no correlation and a positive correlation:

Strongest weakest
0.83, 0.65, 0.03, -0.91

For the McNemar test that was carried out on question 14, participant responses to either one of these sets was categorised as the conception that a positive correlation is stronger than a negative one (figure 8.5). For this question, the findings were not significant (Binomial test, one-tailed test, p = 0.62).

Figure 8.5 McNemar test: question 14 (cases = 12)

		Post test	
		4 (strength misconception)	1 (normal)
Pre test	1 (normal)	0	5
	4 (strength misconception)	3	4

It can be seen from figure 8.5 that three of the participants' responses were categorised as the strength misconception on both the pre test and post test. However, four of the participants' responses were categorised as 'strength misconception' on the pre test, but as normal on the post test. For example, on the pre test one participant indicated on question 14 the set of correlations in which the (strong) negative correlation is weaker than no correlation and a positive correlation, and in explaining this wrote "going from a fairly strong -ve [negative] correlation to zero(ish) then upwards to a +ve [positive] correlation. It's the only one that steadily moves in 1 direction" (O10). In contrast, on the equivalent question 14 on the post test this participant indicated that '-0.91, 0.83, 0.65, 0.03' shows the strongest to weakest relationship and answered "although in opposite directions the strengths are going from left to right" (O10. His emphasis).

Correct and incorrect responses to questions

Participants' whole question scores also provided data concerning correct and incorrect responses to particular questions on both the pre and post tests. Here, participants' responses to questions were categorised as incorrect if:

- They could not answer the question and left the answer blank or wrote "don't know."
- Their response had been categorised as a particular misconception in correlation.
- Their response in the form of an explanation to the question was insufficient.
- An idiosyncratic response was provided by the participant.

In the case of the latter, for example, one participant's response to question 4 was a follows:

> 4. Which correlation coefficient is stronger?
> (a) 0.03
> (b) 0.68 (participant indicated b).
> 4(i) Explain you answer.
> "It shows a coefficient of 68% vs [versus] one of only 3%" (O6).

With regard to the pre and post tests, one tailed McNemar tests were used to see if there were significant changes in the observed frequencies of incorrect and correct responses on questions 2, 6, 3, 7, 10, 4, 8, 11 and 14. That is, all of the questions with the exception of the filler questions and question 13 where only two of the participants' responses changed. There was no significant change(s) in the observed frequencies of responses for questions 2, 6, 3, 7, 10, 8, and 11.

In the case of question 4, there was a significant change in the observed frequencies of responses (Binomial test, one-tailed test, $p = 0.03$) (figure 8.6). From figure 8.6 it can be seen that five participants scored incorrectly on this question in the pre test, but scored correctly on the post test.

Figure 8.6 McNemar test: question 4 (cases = 18)

		Post test	
		1	0
Pre test	0	5	3
	1	10	0

Question 4 concerns the strength of correlation coefficients and asked participants to indicate which of a pair of correlations is stronger. On this question a very low correlation (e.g., 0.03) and a strong positive correlation was provided (e.g., 0.68).

For question 14, there was a significant change in the observed frequencies of responses (Binomial test, one-tailed test, $p = 0.02$). In this question participants were asked to indicate which one of a set of correlation coefficients correctly showed the strongest to weakest relationship and to explain their answer. It can be seen from figure 8.7 that six out of the eighteen participants provided an incorrect response to question 14 on the pre test, but a correct response to question 14 on the post test.

Figure 8.7 McNemar test: correct and incorrect responses to question 14 (cases = 18)

		Post test	
		1(correct)	0 (incorrect)
Pre test	0 (incorrect)	6	7
	1 (correct)	5	0

The above findings relating to question 4 and question 14 suggest that the use of *Link* contributed to participants' understanding of the strength of correlations.

8.3.4 Interactions: case studies

Participants' whole question scores that excluded the filler questions, were examined to see which participant scores had increased from the pre test to the post test. Excluding the filler questions, the maximum possible whole question score for each of the tests was 10. Three out of the eighteen participants' scores had made an increase of 5 scores or more in whole question scores from the pre to the post test (O2, O3, O14). These participants provide case studies that will be described to illustrate their performance on the tests, their misconceptions, and their interactions with the program that might have contributed to their score increase from the pre to the post test. The names of these participants have been changed to assure anonymity.

Case study one

Helen's whole question score on the pre test was 0. Helen's answers to five of the questions on the pre test were confused (questions 2, 4, 6, 7, 13). For example, in response to question 2 she indicated that if a correlational study found a relationship between two variables, you could not ever conclude that there is a causal relationship between the variables. However, in explaining this answer she wrote "if one is the cause of the other there can't be a correlation between them". This is of course not the case. It was difficult to categorise her answer to question 3, where the participant is supposed to interpret a negative correlation from a scatter plot. Here, she commented "higher scores in arithmetic at expense of reading. Each is mutually exclusive, tend to be better at one or the other". It is not clear from these comments whether this participant could or did successfully interpret the pattern on the scatter plot as negative. Helen answered question 4 as follows:

4. Which correlation coefficient is stronger?
(a) 0.01 (Helen indicated a).
(b) 0.64
4 (i) Explain your answer
"1% probability of results being due to chance".

Here, Helen was not clear how to interpret a correlation coefficient and her explanation to her response on this question, suggests that she thought 0.01 represented a level of significance or p (the probability of the observed results or more extreme results occurring by chance alone).

Question 7 asked what correlation coefficient would indicate no relationship between two variables and also asked for an explanation of this answer. Helen did not answer this question appropriately and appeared to provide an example set of variables that would not be related. Her complete answer read as follows:

> "total of time trials and test scores. Total time taken across a series of tests is not relevant, time for each trial would indicate practice effects etc. Total time the trials took is no more than just that - could compare but no useful resulting data".

Question 8 and question 11 asked which one of two correlations is stronger. Although Helen indicated the correct correlation on both these questions, her explanation as to why one correlation is stronger was inappropriate because she remarked that it was the "smallest number". For example, for question 11, Helen responded as follows:

11. Which correlation is stronger?
(a) 0.71
(b) -0.81 (Helen indicated b).
11 (i) Explain your answer
"Smallest number".

To two of the questions, Helen indicated that she did not know the answer. On question 10, which asked whether 0.64 or -0.83 show a correlation, or whether both or neither show a correlation, and also asked to explain the choice, she simply wrote "don't know" twice. Her answer to question 14 was also "don't know". Helen's answer to

question 13, illustrates that she was confused by the meaning of a correlation coefficient and with significance:

13. Which of these shows a correlation?
-0.84
0.02
(a) The first
(b) The second (Helen indicated b).
(c) Both
(d) Neither
13 (i) Explain your choice
"statistically significant".

On the pre test, not one of her responses to any of the questions were categorised as a particular misconception. So, although most of Helen's answers were incorrect, there was no evidence to suggest that she held any particular misconceptions in correlation.

Helen did not experience particular difficulties in completing the first two tasks when using the program. For example, she found out more about the study by selecting the button 'Study details'. When Helen completed activity 1 she read the text on the screen and referred to option 2 'that viewing television violence caused the boys' aggression' and remarked "that one is quite a popular view at the moment". She selected this option and option 4 'that another variable or variables could be responsible for the correlation'. Helen then selected the button 'Done' and when the audio feedback had played provided some thoughts:

"What kind of third variable could be responsible for the relationship that was found to exist between boys' aggression and TV violence?" (Audio feedback to activity 1).

"If their background if their parents are quite easy going and let them get out these 18 films when they're ten ... I'd say it's your parental background really ... and then the peer group as well".

Helen was initially stuck on activity 2 where the participant selects a correlation coefficient from the table that represents the scatter plot that is displayed on the screen. She commented that she was "searching for inspiration" and with reference to the correlation coefficients in the table said "I don't think it is any of the minus ones. Other than that, totally at a loss there". Referring to the target scatter plot she then went on to say "suppose I'll have to say it's 55 ... just because it's evenly matched" and selected the correlation of 0.55. The scatter plot that represents this correlation was displayed on the screen alongside the target scatter plot. Helen then thought out aloud "oh that makes that negative ... that's bringing it all back" and she then selected the correlation -0.07 and then the correlation -0.65.

Having read the text for activity 3, Helen remarked "my initial instinct is to go with the lowest and move upwards ... I know that 64, 55, 65 [0.64, 0.55, -0.65] are significant ... So that must make ... a strong relationship between the variables". She then went on to say "I'm quite put off by -0.07 ... just because it's ... a long way away quite along way the other way that it's quite a strong relationship but it's not necessarily a positive one ... but it's not significant". She then ordered the correlations coefficients as follows:

-0.07 No relationship
0.12
0.18
0.55
0.64
-0.65 Strong relationship

Here, she commented "I'm just going to put them in order numerical order. ... Unless something else comes to me". It can be seen from the above that Helen did, however, arrange the correlations appropriately. Helen went on to look at the additional screen for activity three and here clicked on the various red points on the scale.

Helen's whole question score on the post test was 5 (as opposed to 0 on the pre test). To two questions she wrote "don't know" (questions 2 and 13). In contrast to her answer to question 3 on the pre test, which was scored as incorrect, (and it was not possible to categorise it as a misconception or as normal), her answer to the equivalent question 3 on the post test was correct. Here, she interpreted the pattern on the scatter plot as indicating a negative correlation and provided an exemplary answer:

> "Negative correlation. Those who scored highly on spelling did not necessarily have better memory + [and] those who scored high on memory didn't score well on spelling".

Helen's answers to question 4, question 7, question 8 and question 11 were scored as correct. For example, in contrast to the equivalent question 11 on the pre test, Helen answered question 11 on the post test as follows:

> 11. Which correlation coefficient is stronger?
> (a) 0.73
> (b) -0.84 (Helen indicated b).
> 11 (i) Explain your answer
> "Closer to -1".

On the pre test her answer to question 7 was incorrect, but on this question on the post test she answered the question correctly, providing an appropriate explanation:

> 7. What is a likely value of a correlation coefficient that would tell you that there is not relationship between two variables? (For example, between girls' shoes size and scores on a reading test).
> "0.12"
> 7. (i) Explain your answer
> "Closer to 0 than to 1".

However, as discussed below her answer to question 6 was scored as incorrect, and her answers to question 10 and question 14 indicated that she was not clear about some aspects of correlation. For question 10, Helen answered in the following way:

10. Which of these shows a correlation?
0.68
-0.85
(a) The first (Helen indicated a).
(b) The second
(c) Both
(d) Neither
10. (i) Explain your choice
"Don't know - (was right on program)".

Question 14 asked participants to choose which one of a set of correlations correctly shows the weakest to strongest relationship. Here, in contrast to her answer on this question on the pre test, Helen indicated the correct set of correlations. However, when asked to explain her answer she wrote "furthest from 1 to closest". This is an answer that was assessed to be an insufficient explanation to the question and her answer to question 14 was therefore scored as incorrect.

On the post test, only one of Helen's responses to a question was categorised as a particular misconception. Her response to question 6 was categorised as a causalistic conception of correlation:

6. Suppose there is a correlation of 0.87 between the length of time a person is in prison and the amount of aggression the person displays on a psychological inventory. This means that spending a longer amount of time in prison causes people to become more aggressive. True or false?
(a) False
(b) True (Helen indicated b).
6. (i) Why?
"Closer to 1".

This case study illustrates that *Link* contributed to the participant's general understanding of correlation. Helen's answers to questions on the pre test did not indicate that she held any particular misconceptions about correlation, but that she simply did not know the answers to the questions or was confused by, for example, correlation coefficients and levels of probability. In contrast, her answers to questions on the post

test indicated that she understood aspects of correlation. For example, on question 3 Helen successfully interpreted the pattern on the scatter plot as indicating a negative correlation.

Case study two

Sam's whole question score on the pre test was 1. On seven of the ten questions on the pre test (i.e. excluding the filler questions), Sam answered "don't know" or left the answer space blank.

Sam did not hold a causalistic conception of correlation. His answer to question 2 on the pre test illustrates this:

> 2. If a correlational study finds a relationship between two variables, could you ever conclude that there is a causal relationship between the two variables?
> (a) No (Sam indicated a).
> (b) Yes
> (c) Sometimes
> 2. (i) Why?
> "Because there is never proof of a direct causal connection."

Sam's answer to question 3 was categorised as a unidirectional conception of correlation. Here, a negative correlation is shown on a scatter plot, but Sam's interpretation of the scatter plot was "doesn't show much at all".

When Sam used the program, he found it difficult to interpret the table that is provided on, for example, the introductory screen to the TV violence study. With reference to the correlation coefficients that are displayed in this table he said "these figures mean nothing to me at all" and "I don't even know if a high figure is good or a low figure is good" and "I don't know what total TV means ... or what family income is".

When completing activity 1, Sam read the four options that are provided at the interface in turn, and thought out aloud "it could be all of them". When he had listened to

the audio feedback to this activity he responded:

> "Well, just the fact that correlation studies are meaningless anyway because they don't actually prove anything ... they can only suggest that if you know you eat lots of sausages you like violent programmes ... it's quite a spurious concept for me anyway".

Sam could not understand the question that accompanies activity 2. He remarked "I have no idea what that means", and "how can one figure become a complete graph". With reference to the target scatter plot that indicates a negative correlation, which is displayed on the screen of activity 2, he commented "there's no pattern there at all as far as I can see" and "not much of one anyway". Sam was prompted to select a correlation coefficient in the table and he clicked with the mouse on 0.64. Feedback that read "0.64 is a positive correlation" and a scatter plot that represented this relationship was displayed on the screen. Sam then commented "positive correlation so that goes that way" and "must be a minus one but just in the middle" and "so I'm going to go for double 0 7 [-0.07]". Sam then selected the correlation -0.07 in the table and said "well that's quite close" before he selected the correlation -0.65.

When Sam began working on activity 3, he selected the correlation -0.65 and dragged it to the 'No relationship' position that is provided on the screen of this activity. He, however, then remarked "No, that's wrong" and then selected and dragged all of the correlations to arrange them on the screen as follows:

-0.07 No relationship
-0.65
0.12
0.18
0.55
0.64 Strongest relationship

Sam then selected the button 'Done' in response to the feedback "compare your arrangement of correlation coefficients to the correct arrangement that is shown" and said

"I'm totally confused now". Sam went on to select the button 'Strength' and subsequently clicked on the scale that is provided on the presented screen and read "the nearer to one the stronger the relationship" and commented "oh right even if it's minus one [-1]".

On the post test, Sam's whole questions score was 6. In contrast to his answers to questions 4, 7, 8, 10, 11 where he answered these questions on the pre test as "don't know" or left the answer blank, his answers to these questions on the post test were scored correctly. For example, having answered "don't know" on the pre test for question 7, he answered this question on the post test as follows:

> 7. What is a likely correlation coefficient that you might obtain that would indicate no relationship between two variables? (For example, between girls' shoes size and scores on a reading test).
> "0.0"
> 7. (i) Explain your answer
> "Furthest away from 1".

On the pre test Sam left question 11 blank (he wrote a dash in the answer space on the test), but on the post test he answered this question in the following manner:

> 11. Which correlation coefficient is stronger?
> (a) 0.73
> (b) -0.84 (Sam indicated b).
> 11. (i) Explain your answer.
> "Nearer to -1".

On the pre test Sam left question 14 blank, but on the post test he chose the correct set of correlations that showed the weakest to strongest relationship (weakest 0.04, 0.56, 0.67, -0.79 strongest), but his explanation for this choice was not accurate. He wrote "furthest from 1 to nearest to 1". Here, it is likely that he meant -1 or 1, but it is not certain.

On the post test, Sam's answer to question 6 was categorised as a causalistic conception of correlation. Sam answered this question as follows:

6. Suppose there is a correlation of 0.87 between the length of time a person is in prison and the amount of aggression the person displays on a psychological inventory. This means that spending a longer amount of time in prison causes people to become more aggressive. True or false?
(a) False
(b) True (Sam indicated b).
6. (i) Why?
"High correlation".

This categorisation of his answer was inconsistent because his answer to question 2, which was also designed to identify a causalistic conception, was categorised as normal on the post test:

2. Professor Smith does an experiment and establishes that a correlation exists between variables A and B. Based on this correlation, she asserts that A is the cause of B. Is this assertion correct?
(a) No (Sam indicated a).
(b) Yes
2. (i) Explain.
"There is no causal connection in correlation studies".

Sam's answer to question 3 on the pre test was categorised as a unidirectional conception of correlation and on this question on the post test his answer was categorised in this way again. To this question that asked participants what the scatter plot showed about the relationship between the test scores he wrote "there is no relationship". In fact, the scatter plot indicated a negative relationship.

The above findings suggest that the use of *Link* contributed to Sam's general understanding of correlation. More specifically, Sam answered questions on the pre test by writing "don't know" or leaving the answer blank. However, in contrast to his answers to five particular questions on the pre test that were incorrect, his answers to such questions on the post test were scored as correct. Despite this, Sam's answer to question 3 on the pre test and the post test were both categorised as a unidirectional conception of correlation.

Case study three

Ivor's whole question score on the pre test was 2. On the pre test Ivor's answers to questions 4, 7, 10, 11, 13 and 14 were scored as incorrect because his explanations of the answer were inaccurate and indicated that he was confused by the meaning of correlation coefficients. For example, on question 7 Ivor correctly suggested 0.02 as a likely correlation coefficient that would tell you that there is no relationship between variables, but when asked to explain his answer he wrote "there is only a 2% chance that shoe size/reading test scores are related". (The example variables provided in the question). Similarly, on question 11 he correctly indicated that -0.84 was stronger than 0.73, but explained this response by writing "higher percentage". This explanation is inconsistent with a similar question 8 where Ivor indicated that -0.88 was stronger than 0.02 but justified this answer by writing that "+ and - are not relevant in relation to correlation coefficient strength". This is noteworthy if one is to consider Ivor's answer and inappropriate explanation to question 13:

13. Which of these shows a correlation?
-0.86
0.01
(a) The first (Ivor indicated a)
(b) The second
(c) Both
(d) Neither
13. (i) Explain your choice.
"86% chance that the variables directly affect each other".

Ivor's answer to question 14 was similar in that he provided the correct multiple choice response, but failed to provide an adequate explanation to this response and his answer to the question was therefore scored as incorrect. In this case:

14. Which of the following sets of correlations correctly shows the weakest to the strongest relationship?

Weakest strongest

(a) -0.79, 0.56, 0.67, 0.04

(b) 0.04, -0.79, 0.56, 0.67

(c) 0.04, 0.56, 0.67, -0.79 (Ivor indicated c).

(d) -0.79. 0.04, 0.56, 0.67

14. (i) Explain your answer.

"% increase from lowest to highest".

Ivor did not appear to hold a causalistic conception of correlation: his answer to question 2 was categorised as normal where he explained that "the fact of a correlation in itself does not identify either of the variables as a causal agent". However, it was not possible to categorise his answer to question 6, which was also a question designed to identify a causalistic conception of correlation, as either normal or causalistic because his answer was ambiguous.

On the pre test, Ivor's answer to question 3 was categorised as a unidirectional conception of correlation. With respect to the scatter plot, which indicated a negative correlation, he determined: "there appears to be a better set of results for the spelling test. There is not enough evidence to show that there is a correlation between the two tests. Difference abilities could be being tested".

When Ivor was using the program and was examining the 'Study details' screen, he appeared to have difficulties in making sense of the data table. He thought out aloud "I'm still not getting very much from the table".

When completing activity 1, Ivor considered the available options provided at the interface, or the four possible interpretations of the correlation. He commented that "one or two could be true" or 'that the boys' aggression caused them to watch violence television programmes' or 'that viewing television violence caused the boys' aggression'. However, Ivor only selected the third option 'that the correlation between boys'

aggression and TV violence is spurious' and then selected the button 'Done' and received audio feedback. Ivor then paused and was prompted "what are you thinking?" He then commented:

> "Is it asking me to respond? It obviously is. There could be something around to do with peer group pressure and social and cultural expectations on boys to be tough and so on. So it may be that the violence on television isn't actually causing boys' aggressions levels to be higher. Um it could be social conditioning as much as television programmes".

On activity 2, Ivor reacted to the target scatter plot: "I hate these things", "what do you hate?" (E) "Scatter plots". Ivor paused before he selected any of the correlations in the table that might represent the scatter plot in question and was therefore asked "what are you thinking?" he replied "I haven't got a clue". He selected the correlation 0.64 and then 0.55 and the other correlation coefficients in the table.

On activity 3, Ivor selected and arranged the correlations in the appropriate order. He then selected the button 'Done' and in response to the feedback provided at the interface commented "this is amazing" and "looks like I've put them in the right place". Ivor went on to select the button 'Strength' that invoked the additional screen for activity 3, which he then examined by selecting the various points on the scale and thought out aloud "nearer to one the stronger the relationship. So at least something's stuck in my mind from when I did these things ages ago".

Ivor's whole question score on the post test was 9. On the post test Ivor's answers to questions 4, 7, 10, 11, 13 and 14 were scored as correct. This is in contrast to his performance on the pre test in which his answers to these questions were scored as incorrect. For example, on question 11, Ivor correctly indicated that the correlation -0.81 is stronger than 0.71 and wrote that "the '-' is not relevant to the strength of correlation, but only to the direction". On question 7 Ivor suggested that 0.02 was a likely correlation that you might obtain that would indicate no relationship between two variables, he

explained this by writing "the correlation coefficient is close to 0 and therefore does not support a relationship". In comparison to his answer to question 13 on the pre test as illustrated above, his answer and explanation to this question on the post test were exemplary:

13. Which of these shows a correlation?
-0.84
0.02
(a) The first (Ivor indicated a)
(b) The second
(c) Both
(d) Neither
13. (i) Explain your choice.
"Nearer to -1 than to 0".

On the pre test, Ivor's answer to question 3 had been categorised as a unidirectional conception of correlation. However, his answer to question 3 on the post test was categorised as normal because he described the relationship between the two sets of scores: "the scatter plot suggests that a relationship exists between arithmetic and reading scores. Children strong in one are less strong in the other".

On the post test Ivor's answer to question 6 was not categorised as a causalistic conception of correlation, but as normal because here he did not infer causality. He answered this question as follows:

6. A group of researchers studying the relationship between creative thinking and intelligence administered different measures of creative thinking and intelligence to a sample of high school students. They obtained a correlation coefficient of 0.8 and concluded that high intelligence results in high scores on creative thinking. Is this conclusion warranted from the data?
(a) Yes
(b) No (Ivor indicated b).
6. (i) Explain.
"The explanation could well be turned around. Creative thinking might be the factor behind high intelligence score".

In summary, Ivor's answers to the questions on the pre test indicated that he was generally confused by correlation coefficients and percentages. In addition, his answer to question 3 on the pre test was categorised as unidirectional. However, in contrast to his answers to six questions on the pre test that were incorrect, his answers to these questions on the post test were scored as correct. Moreover, his answer to question 3 on the post test was not categorised as unidirectional, but as normal because he successfully interpreted the negative correlation displayed on the scatter plot.

Summary of case studies

The case studies illustrate that *Link* contributed to students' general understanding of correlation. However, from the case studies it is not possible to determine if the use of the program specifically addressed and changed students' misconceptions in correlation. This is because not one of Helen's answers on the pre test were categorised as a particular misconception. In addition, Sam's and Ivor's answers on the pre test indicated that they lacked the necessary knowledge or were generally confused. However, Sam's and Ivor's answers to question 3 on the pre test indicated that they both held a unidirectional conception of correlation. The findings of these two case studies illustrate that a student's use of the program did not necessarily address a unidirectional conception of correlation: in one case study the participant still held a unidirectional conception having used *Link* (Sam), but in the other case study, the participant no longer held this conception after they had used the program (Ivor).

8.3.5 Outcomes: participants' opinions of the program

The evaluation questionnaire asked participants 'what do you think was the best thing about the program?' Participants' responses to this question included the following:

"It was visually very clear" (O12).

"Easy to use" (O8).

"Explanations provided when required. Easy to use" (O14).

"Easy to work out what to do. Easy to manoeuvre within the program" (O11).

Participants also made comments that concerned the content of the program. For example, one participant wrote that "it explains things well, especially the strength of correlations" (O7). In answering what they thought was the best thing about the program, five of the participants made comments about the answers or feedback that was provided by the program while the participants completed the activities (O1, O9, O14, O15, O18). For example, one participant wrote "the activities have to be completed to obtain the answer (it's too easy to check answers to SAQs [self-assessment questions] without completely committing yourself first)" (O15), and another participant wrote "gave you clear answers immediately after you had responded - so - a quick check for you while still fresh in your mind" (O9). One participant remarked that "you could work through at your own pace" when using the program and more specifically, "you could also learn by comparing mistakes to the correct versions" (O18).

The second question on the evaluation questionnaire asked participants 'what do you think was the worst thing about the program?' In answering this question, one participant wrote "a little unclear on what to do next" (O14). Similarly, another participant remarked "could have used more prompts to advise which screen to use next" (O8) and one participant wrote that "it did not always tell me how to get from (say) activity 2 to the next activity (3)" (O6). One participant commented "I got lost looking for the menu for the next activity. The buttons for moving through the program could be re-worded as I

did not find the current names explicit enough" (O1).

One participant thought that the worst thing about the program was the "lack of infor[mation] about what was actually in the table (even for someone mathematically inclined!)" (O10). Another participant commented that there were "not enough details of [the] study given on the initial screen. Could be expanded slightly to aid overall understanding" (O5).

On the questionnaire, participants were also asked 'what do you think needs changing in the program?' Here, two participants were concerned about the program's usability because one participant wrote "it needs to navigate back without the user having to guess or to use the 'Back' icon" (O6) and the other participant simply wrote "prompts to move on" (O8). Two participants thought the program could be modified by providing more details about the study or the study data. One commented "I feel an opening explanation about the way the figures in the table were obtained may have helped" (O4), and the other participant remarked "more descriptive infor[mation] at the beginning about the purpose behind the study" (O14).

In response to the question 'what did you think of activity 1 in the program?' one participant wrote that "it was quite confusing and I wasn't sure what the purpose of the activity was" (O18). In a similar vein, one participant commented that in activity 1 there was "not enough explanation - too vague" (O7). However, for this question that concerned activity 1, one participant noted "good explanation of why each factor could be relevant" (O11). When asked what they thought of this activity, one participant wrote "thought-provoking. Made you examine other issues that may be involved in this finding and not just accept one explanation" (O5), and another participant simply wrote "thought provoking" (O10).

Participants were asked on the evaluation questionnaire 'what did you think of activity 2 in the program?' One participant thought that it was a "good exercise in

interpreting scatter plots and relating them to coefficients" (O17), and another participant noted "useful to check your understanding of both modes of presenting the figures" (O9). One participant pointed out that although he made an error on this activity "it was very matter-of-fact in its correction and led me easily to the correct answer" (O6). Three participants had concerns about scatter plots: one participant remarked "I don't like scatterboxes! [scatter plots] However, the program felt 'friendly'" (O14), and another wrote that the activity was "marred by my [her] inability to understand scatter plots" (O11). Indeed, one participant noted "had never seen a scatterchart [scatter plot] before so didn't know what I was supposed to be looking for" (O5).

Participants were asked what they thought of activity 3 in the program. Here, participants evidently liked selecting and dragging correlation coefficients across the screen:

> "It interested me that I could move the results. It made me feel I was being tested on my understanding of correlation" (O12).
>
> "I liked the click-and-drag feature, especially being able to reverse a choice easily. maybe it could have mentioned this latter fact" (O6).
>
> "Good - nice to manipulate objects around the screen" (O14).
>
> "I found this the most useful activity" (O18).

8.4 Discussion

Phase two of the formative study had four aims. Findings that pertain to these aims will be discussed in turn.

(a) To investigate whether Link contributed to students' general understanding of correlation

The findings of the second phase of the study showed that there was a significant difference in the pre test and post test means of the overall scores. This was also the case when the filler questions were excluded from the analysis. Additional scores (whole

question scores) on the pre and post tests were calculated for the participants. In terms of these whole question scores, there was also a significant difference between the means of the pre test and the post test. In addition, this significant difference was found if the filler questions were not included in the participants' whole question scores. These findings show that there was a tendency for participants' test scores to increase from pre test to post test.

Participants' whole question scores provided data concerning participants' correct and incorrect answers to particular questions on the pre and post tests. Here, an incorrect answer included, for example, "don't know" and responses that would otherwise be categorised as a misconception (e.g., causalistic). McNemar tests were carried out and it was found that there was no significant change(s) in the frequencies of correct and incorrect responses to questions 2, 6, 3, 7, 10, 8, and 11 on the pre and post tests. However, in terms of the pre and post tests, there was a significant change in the frequencies of correct and incorrect responses on question 4 and question 14. Question 4 concerns the strength of correlations and asked participants which of a pair of correlation coefficients is stronger (e.g., 0.03 or 0.68). Five participants provided an incorrect response to this question on the pre test, but a correct response on the post test. On question 14, six out of the eighteen participants provided an incorrect response on the pre test and a correct response on the post test. It is important to note that this question asked participants to indicate which one of a set of correlations indicates the strongest to weakest relationship and activity 3 of the program involved participants ordering six correlations from that which represents no relationship to that which represents the strongest relationship.

It is clear that participants tended to score better on the post test than on the pre test, but with the exceptions of questions 4 and 14, it is not clear which particular questions were responsible for this improvement. Questions 4 and 14 were designed to test a student's understanding of the strength of correlations and it can be therefore suggested

that *Link* contributed to this particular understanding. Overall, the findings indicate that the program did contribute to participants' general understanding of correlation, but as shall be discussed below, it is not clear whether participants' misconceptions were changed through using the program. In other words, participants' answers on the pre test would indicate that they simply did not know the answer to a question or were generally confused by, for example, correlation coefficients and probability levels, but having used the program they would be able to answer questions on the post test correctly. However, if a student held one of the three misconceptions, which the learner activities in the program were designed to address, then in some cases, the student still held this misconception as indicated by their answers on the post test.

It is possible that the general improvement in the participants' test scores after they had used the program was due to a practice effect. In other words, participants gained practice at answering a particular kind of test by completing a pre test, which meant that they tended to score higher on the post test. This emphasised the need for a control group in the summative evaluation of the developed program. This practice issue aside, there is evidence to suggest that the program contributed to participants' general understanding of correlation.

(b) Find out whether Link affected students' conceptions in correlation

On the pre and post tests, participants' answers to those questions that were designed to identify particular misconceptions in correlation were examined. For particular questions, McNemar tests were used to see if there were significant changes in the frequencies of participants' responses that had been categorised as a particular misconception or as normal. With respect to individual questions, these analyses showed that there were no significant changes in the frequencies of participants' responses on the pre and post tests. In other words, participants' misconceptions did not necessarily change from the pre test to the post test.

(c) Provide a formative evaluation of the program's learner activities and related presentation of topic material.

Although the study was designed to look at learning outcomes and the learning process in relation to using *Link*, findings were outlined that concerned the usability of the program. For example, half of the participants experienced difficulties in the use of the button 'Done' where, for example, they selected it twice when this was not applicable.

The study provided qualitative data concerning participants' interactions with the program. More specifically, the findings suggested ways in which the content of the program could be changed. These changes concern the general design of the program, the study 'TV violence' and the three activities.

It was decided that the general design of the program's human-computer interface should be modified as follows:

- Navigational facilities would be re-designed. For example, a user should be able to move from activity 1 to activity 2 without having to first invoke the TV violence introductory screen.
- Informative feedback would be provided to the user. For example, when a user has been provided with specific feedback that concerns an activity, text should be used to inform a user that they have completed the activity and that they can move on to complete another activity.
- The design of the table that presents data in the program would be modified so that a user can interpret it more easily.

It was clear that activity 1 should be revised for the final program because it was evident from the study that the wording of some of the options could be misinterpreted by students. The feedback to this activity should also be more explicit stating why causality cannot be inferred from a single correlation. This activity should also make use of

examples that challenge a student's causalistic conception of correlation. It was thought that examples would be chosen in which a correlation is established between two variables, but it is obvious why one variable cannot cause the other.

The findings of the study indicated that activity 2 should be changed to provide more detailed feedback concerning the different kinds of relationships that exist between variables. For example, rather than simply informing a user that the target scatter plot represents a correlation coefficient of say -0.65 and stating that this is a negative correlation, feedback provided to the student should describe a negative relationship and compare it directly to no relationship and a positive relationship. In addition, students might not know that a correlation coefficient is represented by a scatter plot and the final program should make this explicit.

The feedback to activity 3 should be changed to clearly state that when assessing the strength of correlation coefficients, the size rather than the direction of a relationship is what is important. The screen that provides a scale to indicate the strengths of correlations should not be used in the third version of the program. Instead, the feedback to activity 3 should be revised to provide more detail about the strength of correlations.

The first and second prototypes of *Link* present some correlation coefficients as statistically significant. However, students might interpret a significant correlation differently from one that is simply reported as a correlation. Hawkins and her colleagues have pointed out that students might not interpret the term significant as meaning "statistically significant", but think that a significant statistic means that it is important or of "practical significance" (Hawkins et al, 1992, p. 87). It was decided that the final version of *Link* should not present particular correlations as statistically significant because it was necessary to focus on students' understanding of correlations in the summative study. Inevitably students will hold confusions concerning statistical significance, but this would require further research. This issue is briefly re-considered when the final design of *Link* is presented in chapter 9.

(d) Pilot tests that were designed to provide an assessment of students' understanding of correlation.

Two equivalent tests were developed to be used in the second phase of the formative study. An analysis of participants' performance on these tests suggested that some questions were ambiguous in their wording and could have produced participant responses that were not easy to categorise (e.g., test A: question 6). However, a number of the questions were valuable in that they identified participants' conceptions in correlation (e.g., question 3 and question 14 on test A and test B). Both of the tests took the participants at least twenty minutes to complete, which together with working through the program, meant that a session lasted for about forty minutes. With this in mind, it was decided that is was not necessary to include the filler questions in the tests used in the summative evaluation study. For the summative evaluation of *Link* the tests were modified: filler questions were not included in the revised tests in correlation and those questions that were ambiguous were re-worded for these tests. (See chapter 9).

Questions on the tests did identify participants' misconceptions in correlation. It was evident from participant responses on either the pre test or the post test that:

- Thirteen out of the eighteen participants held a causalistic conception of correlation (O2, O3, O5, O6, O7, O8, O10, O12, O13, O14, O15, O16, O17).
- Six out of the eighteen participants held a unidirectional conception of correlation (O3, O5, O8, O11, O12, O14).
- Nine out of the eighteen participants held the conception that a positive correlation is stronger than a negative correlation when this is not the case (O4, O5, O8, O10, O13, O15, O16, O17, O18).

However, not every question that was designed to pick up on a particular misconception in correlation did so. For example, both question 3 and question 7 were designed to identify a unidirectional conception of correlation, but one participant's answer to

questions 3 was categorised as a unidirectional response and on question 7 their answer was categorised as a normal response. It was also found that although question 2 and question 6 were both designed to identify a causalistic conception of correlation, the pattern of responses to these questions was different. For example, in the case of question 6 more of the participants' answers were categorised as causalistic or were not categorised as either normal or causalistic, than in the case of question 2. This means either that participants' answers to particular questions have to be reviewed to see what type of question elicits a particular misconception, or that participants themselves do not necessarily hold a conception that is consistent. Both of these possibilities are likely. However, question 6 on test A was thought to be ambiguous and was re-worded for subsequent research. (See chapter 9).

8.5 Expert evaluation of the program

Three experts at the Open University took part in this evaluation of the second prototype. One of these experts was an experienced software designer who had an academic background in statistics, one had specialist knowledge of psychology and statistics, and one specialised in the field of computers and learning. These experts were asked to evaluate the prototype and were told that it was a revision program for correlation that was designed to be used by students taking a degree in psychology. They were also informed that one of the studies 'TV violence' had been implemented in the prototype and to evaluate the prototype, to select the button 'TV violence' on the introductory screen. They completed an adapted version of the program evaluation questionnaire that was used in the second phase of the formative study.

In response to the question, 'what do you think was the best thing about the program?' one of the experts wrote "I liked the voice explanations and the interactive nature of the program" (E3), and one commented "the graphical display of correlations" (E2). The experts all raised concerns about the program's navigation. For example, when asked 'what do you think was the worst thing about the program?' one expert noted

"navigation - knowing when an activity finished" (E1), and another wrote "some rather odd navigational movements" (E2). When asked what they thought needed changing in the program, one expert said "I would change the menu bar at the bottom" (E3) and similarly, one expert thought the "lower row of navigation buttons" needed changing (E2).

Two of the three experts commented on how the program could use individualised instruction for some of the activities. One expert wrote that he "would also like more individualised instruction" (E3). When asked what they thought about activity 1, two experts also raised this design issue:

> "I particularly liked the voice explanation. I would have liked more individual feedback. For example, if I selected 1 & 2 [options] to only hear explanations for 3 & 4 [options]" (E3).
>
> "[Program] reaction is non-contingent on user's responses" (E2).

When asked what the experts thought of activity 2 in the program, one commented "a nice display of the various correlations", but that "the correct answer is not strongly flagged" (E2). One expert was concerned that only general canned feedback was provided to a user's answer to activity 3: "computer reaction is again non-contingent on user's response" (E2).

The questionnaire asked 'what do you think were the best instructional strategies that were provided in the program?' Here, one expert simply wrote "audio feedback. Dragging the coeffs [coefficients]" (E1), and one expert commented "showing graphs of the correlations (best of all)" (E2). The experts were also asked what they thought were the worst instructional strategies provided by the program. Here, one re-iterated that he "would have like[d] more individualised instruction" (E3), and another felt that "activity 1 worked least well" (E2).

The experts were asked what kinds of instructional strategies should be provided in

the program. One expert thought that more examples in the form of scatter plots could be provided in activity 2 (E1), and one expert raised the following question: "does there need to be [a] more explicit demonstration of the strength of negative correlation?" (E2).

8.6 Summary

This chapter described the second phase of the formative evaluation study, which indicated that the second prototype of *Link* contributed to students' general understanding of correlation. Indeed, there was a significant increase in mean scores from the pre test to the post test. Although the findings did not indicate that the program affected students' misconceptions concerning correlation, the prototype of *Link* only provided three activities that were designed to address particular misconceptions. The findings of the formative and expert evaluations indicated that further development of *Link* was required.

In the formative study tests in correlation were piloted so that they could be used in a summative evaluation of *Link*. The formative evaluation provided valuable qualitative data relating to the learning process and case studies were used to illustrate this. In addition, qualitative data concerning students' interactions with the program and the variety of student responses to the program's learner activities were described.

An expert evaluation of the program was also described in this chapter. The findings of the formative and expert evaluations were used to inform the final design of *Link*, outlined in the following chapter, in which the summative evaluation of *Link* is also reported.

Chapter 9

A summative evaluation study of *Link*

9.1 Introduction

This chapter outlines the final version of *Link*, which was developed in response to the findings of the formative study and the expert evaluation that were described in the previous chapter. This chapter also describes a summative evaluation study of the developed program. The primary aims of this research were to investigate whether *Link* contributed to students' general understanding of correlation and whether the learner activities affected students' misconceptions.

9.2 Final design of *Link*

Following the second phase of the formative evaluation, the final version of *Link* was developed. To improve the usability of the program, the navigational facilities were re-designed. *Link* was also improved by including the following:

- An 'OK' button is used for each learner activity. This can be selected by a user to obtain feedback.
- Feedback is provided for when a user has completed an activity, to inform them that they can move on to attempt another one.

The data tables that provide variables and correlation coefficients were also re-designed. As considered below, in the final program, a user can select a pair of variables from a data table and details of how the variables were measured in a study is provided. If a user selects the term 'Correlations' in a table, they are provided with details of this measure.

With the summative evaluation study in mind, *Link* was designed so that when a student used the program, a user log was created. The user log records the buttons that a

user selects, the screens that they visit, a user's responses to the learner activities and the feedback they receive for each activity. Appendix C contains sample program code from *Link* and appendix F contains details of the program's design including the outline of each study, the text for the activities and feedback to the activities.

The introductory screen that outlines the primary objectives of the program was maintained, but the buttons that can be selected by a user to choose a study were changed. In the final version of *Link*, a user can select the study 'Health events' or the study 'Infant engagement' on the introductory screen because, as described below, the final program provides two studies from the psychological research literature.

In chapter 4, it was proposed that the acquisition of statistical concepts is facilitated if they are presented to psychology students in the context of a psychological study. Drawing on theoretical perspectives (e.g., Bransford et al, 1990), it can be argued that this context should be realistic and meaningful to the learner. It was clear from the review of computer-assisted learning programs that was described in chapter 4 that existing programs do not use real data from psychological research. It was therefore decided that the final version of *Link* should use empirical studies from the psychological research literature, but if this was the case, then the data sets from these studies would be required for the program. A thorough search for available data sets from psychological studies was conducted. However, this search for studies and associated data sets was seriously constrained by the fact that a study would have to provide a variety of variables that could be correlated to give a set of correlations that could be used in *Link*. Specifically, bivariate data was required that provided positive, negative and weak correlations. Text and web-based resources that provide data sets were consulted for possible use in the program (e.g., Hand, Daly, Lunn, McConway & Ostrowski, 1994). However, these resources could not be used because they do not necessarily provide data that is suitable for correlation and/or do not provide data that were obtained from psychological studies, and in the case of web-based resources, permission must often be given by the researcher to

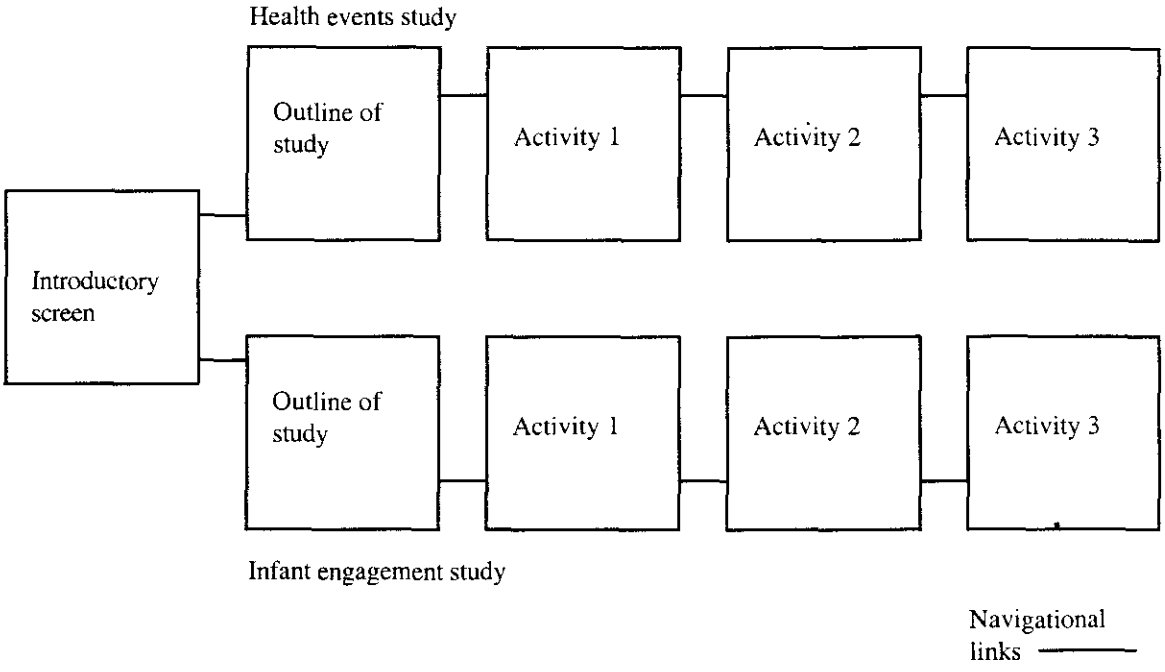
use his or her data sets.

Contact was made with two researchers at the Open University who had each carried out research in psychology and who were willing for brief descriptions of their study and the associated data to be used in *Link* (Cohen & Java, 1995; Oates, 1998). Permission was granted by each researcher to use their studies in the program and a summary of each of the studies was written to provide an outline of each in the program. Permission was also given to use the data sets from these studies to obtain the correlation coefficients and scatter plots that are used in *Link*. One of the studies concerns people's memory for medical history (Cohen & Java, 1995) and the other relates to infant engagement and maternal variables (Oates, 1998). For *Link*, the former study was called the Health events study and the latter was called the Infant engagement study. The program contains two sections, each of which provide:

- A screen that presents a brief outline of the study. This includes a description of the variables.
- A table of data containing correlation coefficients.
- Three learner activities that use data from the study.

The revised schema for the developed program is shown in figure 9.1. The navigational links shown on this schema are not the only possible links available to a student while they use the program. For example, a user can move from activity 3 in the Health events section to the screen that outlines the Health event study.

Figure 9.1 Schema for Link



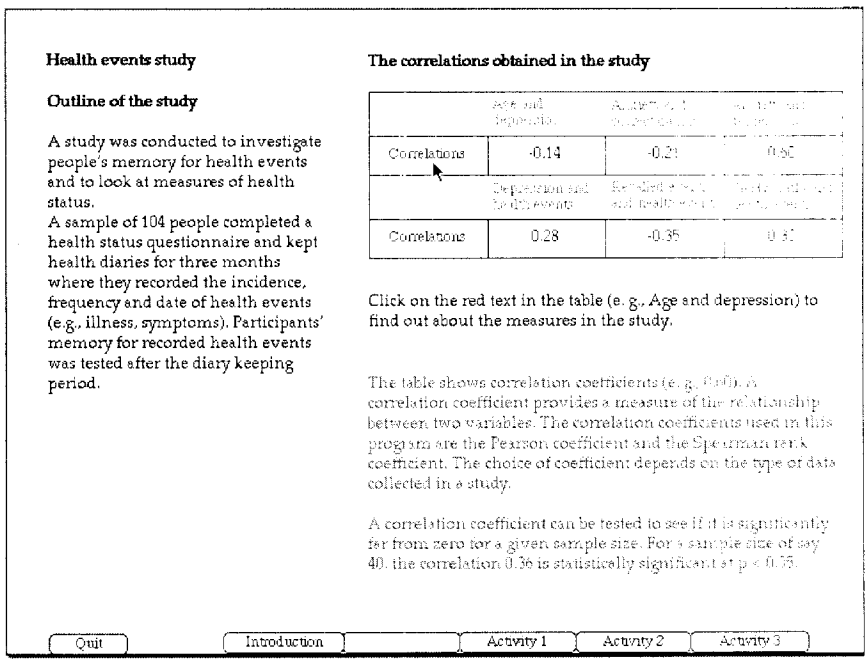
The earlier prototypes of *Link* provide correlations that are reported to be significant. However, the final version of *Link* does not present correlations that are reported as statistically significant or describe statistical significance in any detail. This is because students hold misconceptions relating to significance and probability, and research on students' learning of probability and statistical inference was beyond the scope of the work described in this thesis. Nevertheless, as described in chapter 2, the testing of a correlation to see if it is significant is usually covered in the curriculum. In *Link*, therefore, the correlation coefficient as a measure of a relationship between two variables is described and it is stated that a correlation can be tested to see if it significant.

As noted above, each of the studies in the program has a screen that provides an outline of the study. Here, as shown in figure 9.2, a user can select the term correlations from the data table and text at the interface informs a user that a correlation coefficient is a measure of a relationship between variables, and that:

The correlation coefficients used in this program are the Pearson coefficient and the Spearman rank coefficient. The choice of coefficient depends on the type of data collected in a study.

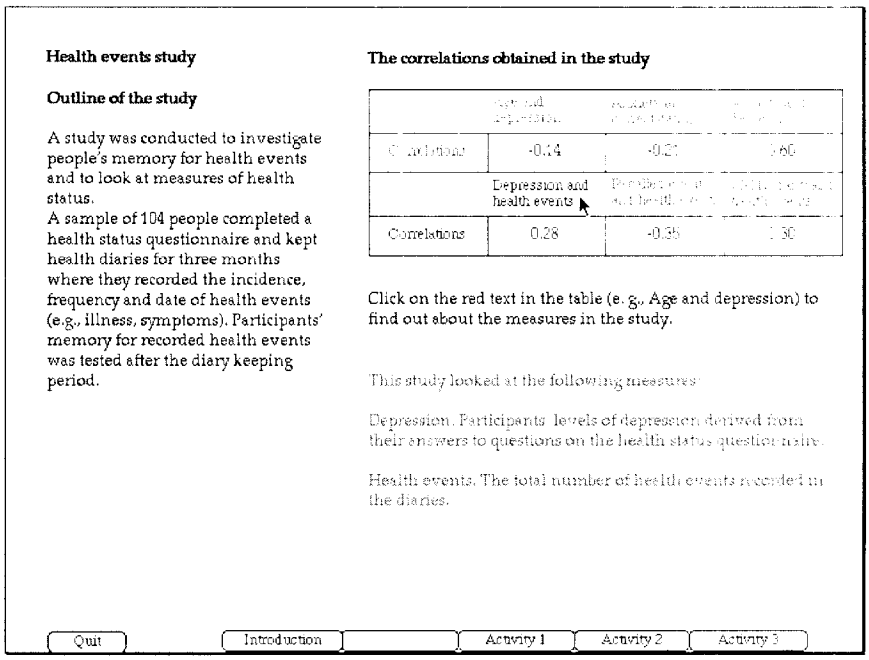
A correlation coefficient can be tested to see if it is significantly far from zero for a given sample size. For a sample size of say 40, the correlation 0.36 is statistically significant at $p < 0.05$.

Figure 9.2 Health events study



In using *Link*, a student can initially choose to select the study 'Health events' or 'Infant engagement' from the introductory screen. If, for example, a user selected the Health events study they would be provided with a screen that outlines the study, the variables studied and the correlations obtained. On the data table provided on this screen, a user can select the term correlations to find out about the coefficient as a measure, or they can select pairs of variables in the table in turn. For instance, if a user was to select 'Depression and health events', text is provided that describes how these variables were measured in the study (figure 9.3).

Figure 9.3 Health events study. A pair of variables has been selected



The three kinds of learner activities that were implemented in the earlier prototypes of *Link* were modified with regard to the findings of the formative study. For example, the question and options for activity 1 were re-worded because they were not clear to students who took part in the study. The expert evaluation also informed the development of *Link* because, for example, activity 3 were revised to provide feedback that is conditional on a user's response to the activity.

9.2.1 Activity 1

In *Link*, two learner activities were implemented that were designed to address a causalistic conception of correlation. One of these activities is provided in the Health events study section of the program and the other in the section of the program that provides the infant engagement study. They are comparable activities and are therefore both called activity 1.

The basic design of activity 1 was maintained from the second prototype, but the question and options for this activity were re-worded. This activity provides feedback in

the form of text and audio, which was based on the feedback designed for the earlier prototypes. However, the feedback is contingent on a user's response to the activity where they must interpret a correlation (figure 9.4). The program was designed so that if a user selects particular options for the activity they are provided with feedback that relates to the interpretations of a correlation that they did not select (appendix C). The feedback emphasises why causality cannot be inferred from a single correlation and uses an additional example in which a correlation has been obtained between two variables, where it is clear that one variable cannot cause the other. In the Health events section of the program, feedback to activity 1 briefly outlines a correlation that was obtained between brain size and IQ (Willerman et al, 1991) (figure 9.5). As depicted in figure 9.6, feedback to activity 1 for the Infant engagement study reports on a correlation that was obtained between the age at which babies start to crawl and monthly temperature (Benson, 1993).

Figure 9.4 Activity 1

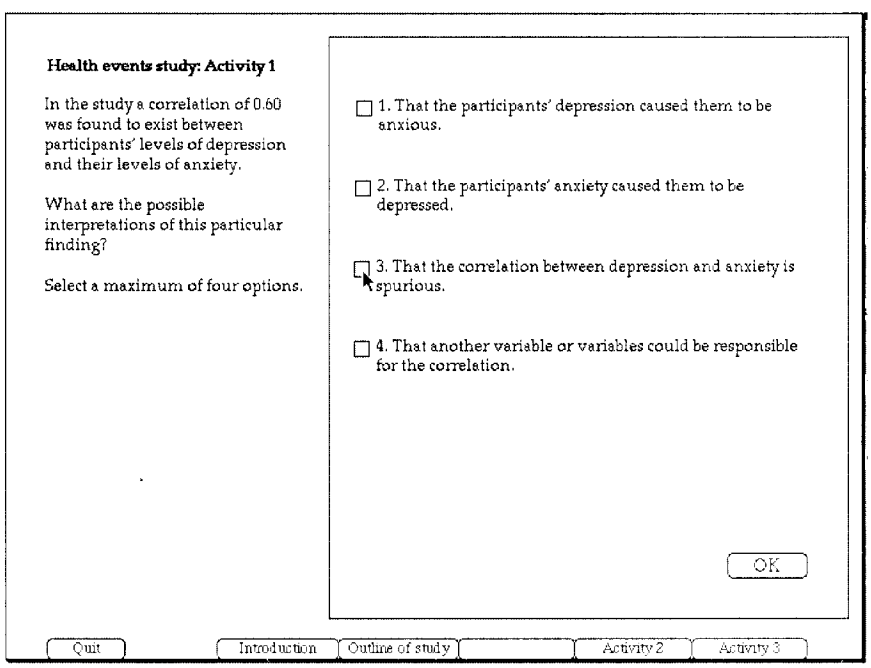


Figure 9.5 Health events study. Feedback to activity 1

Health events study: Activity 1

Brain size and IQ

One study found a positive correlation between a measure of brain size and IQ scores (Willerman et al, 1991). This does not of course mean that brain size is the cause of IQ, or vice versa. The correlation could be spurious, or other variables could be responsible for the correlation.

Reference
Willerman et al (1991) In Vivo Brain Size and Intelligence. Intelligence, 15, 223 - 228.

You can now move on to another activity.

☐ 1. That the participants' depression caused them to be anxious.

☐ 2. That the participants' anxiety caused them to be depressed.

☒ 3. That the correlation between depression and anxiety is spurious.

☒ 4. That another variable or variables could be responsible for the correlation.

Quit | Introduction | Outline of study | | Activity 2 | Activity 3

Figure 9.6 Infant engagement study. Feedback to activity 1

Infant engagement study: Activity 1

Babies crawling and temperature

One study found a negative correlation between the average age at which babies first start to crawl and average monthly temperature for the sixth month following birth (Benson, 1993). This does not of course mean that low monthly temperatures cause babies to crawl late. The correlation could be spurious, or other variables could be responsible for the correlation.

Benson, J. B. (1993) Season of Birth and Onset of Locomotion: Theoretical and Methodological Implications. Infant Behavior and Development, 16, 69 - 81.

You can now move on to another activity.

☒ 1. Infant age is causally related to levels of infant engagement.

☒ 2. The correlation between infant age and levels of infant engagement is spurious.

☐ 3. Another variable or variables could be responsible for the correlation between infant age and level of infant engagement in the experiment.

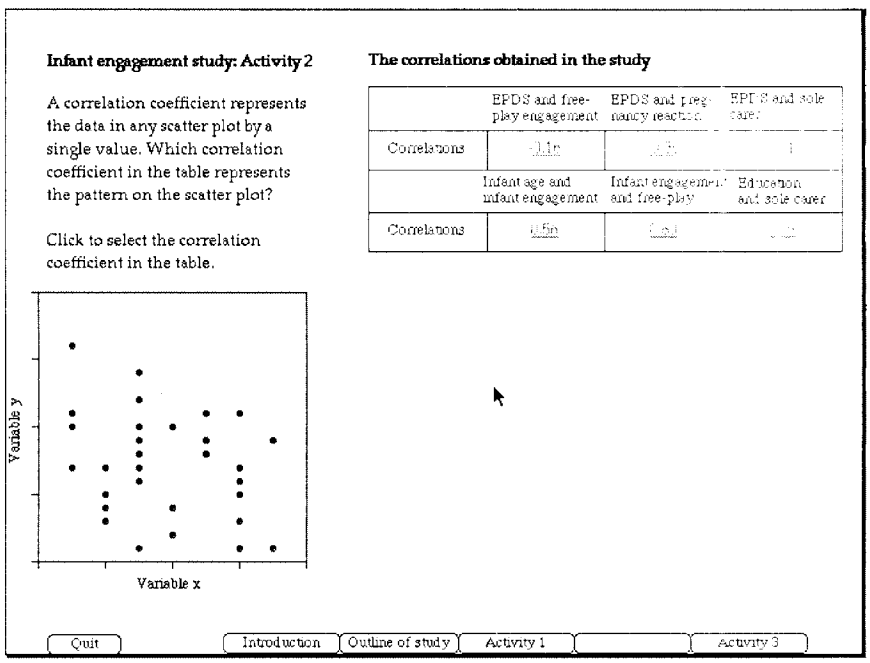
Quit | Introduction | Outline of study | | Activity 2 | Activity 3

9.2.2 Activity 2

In *Link*, two learner activities were included that were designed to address a unidirectional conception of correlation. One of these activities is in the section of the program that provides the Health events study and the other in the section of the program that provides the Infant engagement study. They are comparable activities and are therefore both called activity 2.

The basic design of this learner activity was maintained from the second prototype: in the learner activity a student must decide from a set of correlations the correlation that represents the scatter plot that is displayed. However, text for this activity was added to state that a correlation represents the data in any scatter plot by a single value (figure 9.7).

Figure 9.7 Infant engagement study. Activity 2



This activity was designed to address the unidirectional conception of correlation because the target scatter plot represents a negative correlation. Each of the correlations in the table are linked to a scatter plot so that when a user selects a correlation a scatter plot that represents the relationship is presented alongside the target scatter plot. A user is

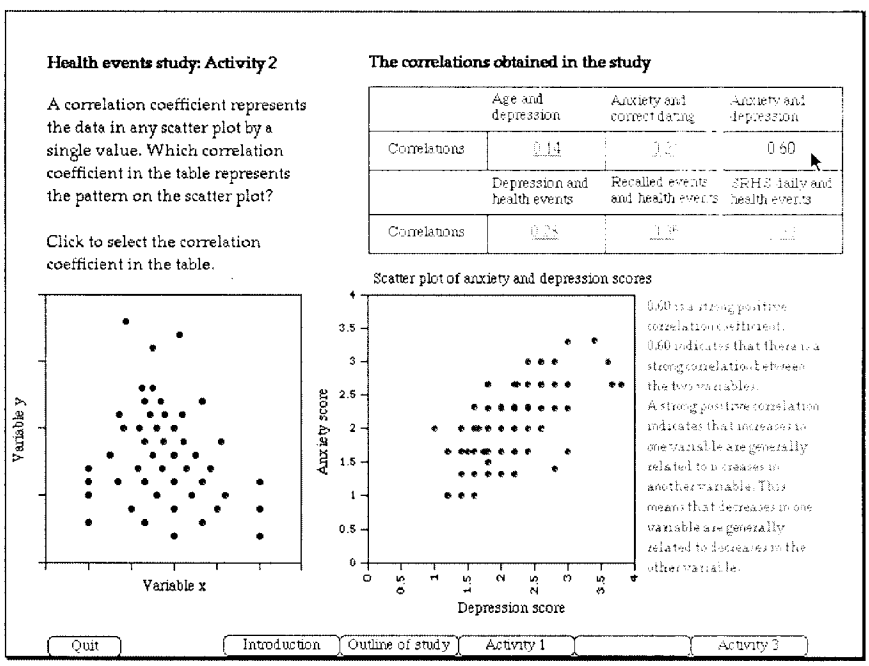
therefore provided with graphical feedback in the form of a scatter plot and text-based feedback that:

- States what kind of relationship is represented by the correlation coefficient selected by the user (e.g., negative correlation).
- Describes the relationship represented by the correlation coefficient selected.

So, if a user selects the correlation 0.60 in the table they are provided with the scatter plot that represents this relationship and appropriate feedback that states:

> 0.60 is a strong positive correlation coefficient. 0.60 indicates that there is a strong relationship between the two variables. A strong positive correlation indicates that increases in one variable are generally related to increases in another variable. This means that decreases in one variable are generally related to decreases in the other variable.
>
> (figure 9.8).

Figure 9.8 Health events study. Activity 2



When a user selects the correlation -0.35, they are provided with the following feedback:

> -0.35 is a moderate negative correlation coefficient that represents the pattern on the scatter plot. -0.35 indicates that there is a moderate correlation between the two variables
>
> (excerpt from feedback to activity 3 for Health events study. Appendix F).

Possible feedback to activity 2 uses the terms strong, moderate, weak and very weak to describe the different correlations. These terms are based on Coolican (1990, p. 210).

9.2.3 Activity 3

Two learner activities in *Link* were included that were designed to address the conception that a positive correlation is stronger than a negative correlation when this is not the case. One of these activities is in the section of the program that provides the Health events study and the other is in the section of the program that provides the Infant engagement study. They are comparable activities and are therefore both called activity 3.

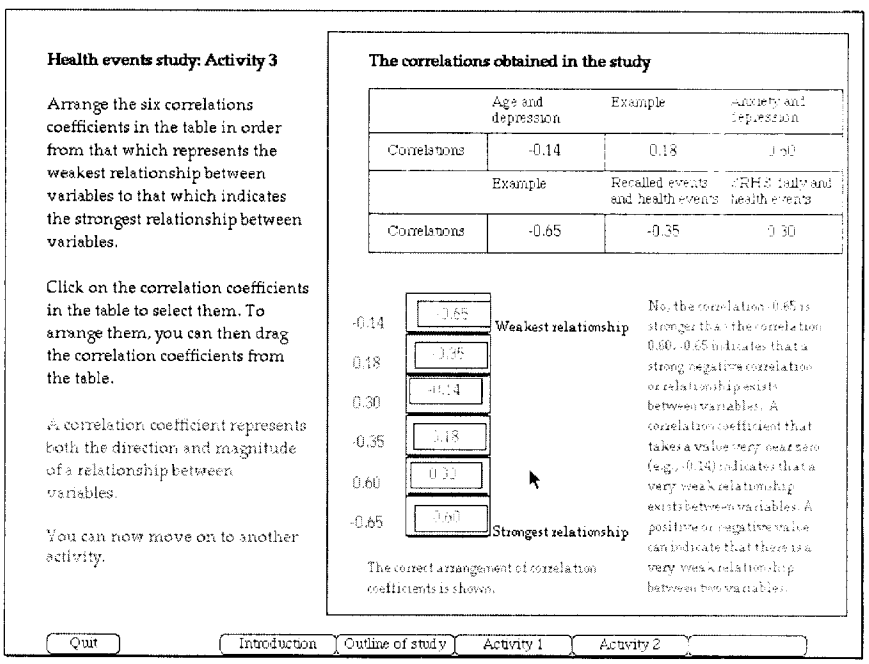
The basic design of this activity was maintained from the second prototype: a user arranges a set of six correlation coefficients from that which indicates the strongest relationship to that which indicates the weakest relationship. However, for the final version of *Link* the 'no relationship' position was changed to read the 'weakest relationship'. This change was made because a statistical advisor pointed out that no relationship is represented by a correlation of zero. Although the activities use real data from the two studies, in the case of activity 3, example correlations had to be used for the set of six correlations that are displayed in the table of data. Specifically, two example correlations of -0.65 and 0.18 replaced two of the six correlations provided for the Health events study and two example correlations of 0.32 and 0.26 replaced two of the six correlations provided for the Infant engagement study. This was necessary for the design of activity 3 because a set of correlations was needed that provided correlations representing weak, positive and negative relationships, but that also provided a negative

correlation that was the strongest correlation of the set.

Feedback to activity 3 is appropriate to a user's arrangement of correlations. Three kinds of feedback can be provided depending on whether a user has correctly arranged the correlations, or has positioned 0.60 as the strongest relationship, or has incorrectly positioned the correlations in an unspecified arrangement. For example, if a user places the correlation 0.60 rather than -0.65 in the 'strongest relationship' position, feedback informs the user that 'the correlation -0.65 is stronger than the correlation 0.60' (figure 9.9). Appendix C provides sample program code for this set up (see example C.8) and appendix F details the possible feedback to activity 3.

This activity also provides feedback in the form of an arrangement of correlations and text-based feedback that makes it clear that when assessing the strength of correlation coefficients, it is important to consider the size and the direction of a relationship (figure 9.9).

Figure 9.9 Health events study. Feedback to activity 3



9.3 Introduction to summative evaluation

The primary purpose of the study that is described in this chapter was to provide a summative evaluation of the developed program, *Link*. The evaluative framework that was outlined in chapter 5 was employed in this study (table 5.3). A formative evaluation of an earlier version of the program suggested that *Link* significantly contributed to students' understanding of correlation, but did not use any forms of control. In chapter 5 it was emphasised that a summative evaluation of *Link* should use two forms of control: a basic control and an instructional control. The basic control group of students completed a pre test and then worked through a section of a computer-assisted learning program that covered the psychology of associative learning before they completed a post test. The instructional control group of students also completed a pre test, but worked through paper-based instructional materials that covered the topic of correlation prior to completing the post test.

Delayed post tests were not employed in this evaluation study because it was thought that students might have received additional forms of instruction that would have had implications for their performance on a test that assessed their understanding of correlation.

The primary aims of the summative evaluation are summarised as follows:

(i) To investigate whether *Link* contributed to students' understanding of correlation.

(ii) To find out whether learner activities in the computer-assisted learning program affected students' misconceptions concerning correlation.

9.4 Method

9.4.1 Design

The study used a pre-test-post-test control group design involving psychology students at the University of Luton.

9.4.2 Participants

Fifty students (thirty eight females and twelve males) who were studying psychology at the University of Luton took part in the study. Forty of these students were in the second year of their undergraduate degree programmes and ten were in their final year. Thirty eight of the students were taking psychology as a single honours degree and twelve of the students were taking psychology as part of a combined or joint degree. The students were paid £4 for their participation. The mean age of this group of students was 25 years (S.D. = 7.79, minimum = 19, maximum = 48). Thirty two out of the fifty students had a GCSE in Mathematics, nine students had an O' level in Mathematics and nine students had neither of these qualifications. Out of the nine students who had neither qualifications, three students had completed their secondary education in Greece or Italy, two students had a CSE in mathematics and one student had an Irish Leaving Certificate in mathematics. The students who were in the second year of their degree programme had completed the course Research and Experimentation I which covered correlations and relationships between variables, and the third year students had completed this course and Research and Experimentation II which covered correlation and multiple regression.

With regard to computing experience, the average length of time for having used computers was five years (mean = 5.25, minimum = 1.5 years, maximum = 15 years). Thirty three of the participants had used both an Apple Macintosh and an IBM PC-compatible and seventeen had only used an Apple Macintosh. All of the fifty participants reported that they used Microsoft Word and that they made fairly regular use of the

computer: thirteen participants indicated that they used it every day, twenty one participants used it every two to three days, eleven participants used it once a week, three participants used it once a month and two participants used it less than once a month.

Prior to recruitment, participants were randomly assigned either to the *Link* group or to the basic control group. It was planned that a lecture class of students were to act as the instructional control group. However, only six students were recruited in this class for the instructional control. Subsequently, students were recruited and assigned to the instructional control group, the basic control group or the *Link* group. There were 17 participants in the *Link* group, 17 participants in the basic control group and the instructional control group consisted of 16 participants. Descriptive statistics for these three groups are summarised in table 9.1. It can be seen from this table that, in general, the participant descriptive statistics for the three groups are comparable. There are, however, exceptions. From table 9.1 it is clear that although the gender mix is roughly equivalent in the control group and the instructional control group, compared to these two groups, the *Link* group had a smaller proportion of females and a higher proportion of males. It is also noteworthy that the instructional control group had a higher proportion of second year students and a lower proportion of final year students than the *Link* and control groups. With regard to mathematics qualifications, the *Link* group and the control are equivalent, but compared to these groups the instructional control group has a greater proportion of students who had obtained a GCSE in mathematics and a lower proportion of students who have neither an O' level or GCSE in mathematics.

Table 9.1 Participant descriptive statistics for the three groups

		Link n = 17	Control n = 17	Instructional control n = 16
Age (years)	mean	24	25	25
	S.D	7.73	7.53	8.50
Gender	females	65% (11) [a]	82% (14)	81% (13)
	males	35% (6)	18% (3)	19% (3)
Year	Second	76% (13)	71% (12)	94% (15)
	Final	24% (4)	29% (5)	6% (1)
Degree	Single	76% (13)	65% (11)	88% (14)
	Combined/ joint	24% (4)	35% (6)	12% (2)
Maths qualification	GCSE	59% (10)	59% (10)	75% (12)
	O' level	18% (3)	18% (3)	19% (3)
	neither	23% (4)	23% (4)	6% (1)
Computer use (years used)	mean	6	5	4
	S.D	4.51	2.98	3.81

[a] Frequencies are given in brackets.

9.4.3 Materials

Hardware and software

The developed program, *Link*, was stored and run on an Apple Power Macintosh 8100/80 AV. A user log was created when each of the participants used *Link*. This log was in the form of a text file that detailed the relevant participant's actions at the human-computer interface and the feedback they received. A sample user log is provided in appendix G.

Sections of a computer-assisted learning program, *The Secrets of Psychology - Associative Learning* (Bond, The Flinders University of South Australia) were also stored and run on the Power Macintosh described above. This program is a HyperCard application that consists of nine chapters that cover the psychology of associative learning. Students completed chapter 5 of this application that covers the topic operant conditioning and which provides text, audio and video elements related to this topic.

Paper-based instructional materials

Paper-based instructional materials that covered the topic of correlation were developed. These materials were adapted from a text that provided text exposition, scatter plots and related graphics (Coolican, 1990). The materials provide text and scatter plots relating to the concepts of positive, negative and zero correlations, the strength of correlations and causality. The materials also included three self-assessment questions for the students to answer by writing in spaces provided, and answers to these questions at the end of the materials. The self-assessment questions and associated answers were adapted from course materials (Open University, 1990).

Participant profile

This questionnaire was adapted from the participant profile used in the formative evaluation study, and was completed by participants to collect data concerning, for example, the participant's gender, age and qualifications.

General instructions for Link group

This outlined the purpose of the session and told the participant what would happen during the session. Participants were asked to work though the program by completing a set of tasks. Firstly, participants were asked to select the 'Health events' study, find out about the study and to complete in turn, activity 1, activity 2 and activity 3 for this study. Secondly, participants were asked to select the 'Infant engagement' study, find out about

the study and complete activity 1, activity 2 and activity 3 for this second study.

General instructions for basic control group

This outlined the purpose of the session and told the participant what would happen during the session. Participants were asked to complete specified sections of the chapter Operant Conditioning: Reinforcement from *Secrets of Psychology - Associative Learning*.

General instructions for instructional control group

This outlined the purpose of the session, what would happen during the session and that as participants they were to complete paper-based instructional materials by reading through all the material and by answering the self-assessment questions. Participants were asked to answer each question by writing in the space provided and they were informed that they could refer to the answers to the questions that were given at the end of the materials.

Additional instructions

These additional instructions were read by participants who were assigned to act as case studies in the investigation. These instructions told the participants that they were to be observed while they used the program and were asked to think aloud while they did this.

Audio cassette recorder

When participants, who were assigned to act as case studies, used *Link* their think aloud was audio recorded.

Tests in correlation

Two equivalent tests in correlation were used in the study (test C and test D). These tests and the associated scoring schemes were based on versions that were used in the formative evaluation study. Test A and test B that were used in the formative study were modified to give test C and test D. Specifically, filler questions in tests A and B were excluded, question 6 on test A was re-worded, and question 10 and question 13 on tests A and B were re-worded.

With the exception of question 2 on test C and test D, each of the questions had two parts. For the first part of a question (e.g., question 1) the participant had to give a short answer to the question or they had to indicate one of a set of available answers given on the test, and for the second part of a question (e.g., question 1i), the participant was asked to provide an explanation for their answer. A score of 1 was given for a correct answer to a question and a score of 1 was given for an appropriate explanation of this answer. This scoring provided the participant's overall score on a test where the maximum score for each test was 19.

Questions on these tests had been developed to identify students' misconceptions in correlation. Each of the equivalent tests in correlation included the following:

- Two questions (question 1 and question 4) designed to identify a causalistic conception.
- Four questions (question 2, question 5, question 7 and question 9) designed to identify a unidirectional conception.
- Four questions (question 3, question 6, question 8 and question 10) designed to test students' understanding of the strength of correlations.

To categorise students' answers to the above questions, a participant's responses to a two part question was coded as appropriate if the question was answered correctly and an

appropriate explanation of the answer was also provided. By using the devised marking schemes, participants' answers could be categorised as particular misconceptions. Participants' answers to the two part questions were therefore categorised as an appropriate conception or as a misconception (e.g., causalistic conception).

By using the devised scoring schemes for the tests in correlation, a second researcher at the Open University scored a sample of the pre and post tests and checked the student responses on these tests that had been categorised as misconceptions. (See 9.5.2 Outcomes).

9.4.4 Procedure

Participants in the *Link* and control groups worked individually in a cubicle in the psychology laboratory at the University of Luton. At the beginning of the session, the participants were provided with relevant instructions concerning the session. During the session, participants completed two equivalent tests in correlation. The order of the tests was randomly assigned so that participants either completed test C prior to using the computer and completed test D after they had used either program, or vice versa. Prior to using *The Secrets of Psychology - Associative Learning*, participants in the control group were briefly shown how to use the human-computer interface of this computer-assisted learning program. Participants in the *Link* group were not observed while they used the program, but the program provided a user log of relevant participant actions at the human-computer interface and the feedback received. At the end of the session participants completed a test in correlation and were de-briefed.

Six of the participants in the instructional control group worked individually, but in parallel in a lecture theatre at the University under test conditions. Two participants in the instructional control group worked individually in a psychology laboratory and eight of the participants completed the materials individually, but worked in parallel with one other participant in the laboratory. At the beginning of the session, the participants were

provided with the relevant instructions concerning the session and during the session, participants completed two equivalent tests in correlation where test order was randomly assigned so that participants either completed test C prior to completing the instructional materials and completed test D after they had worked through the materials, or vice versa. At the end of the session participants were de-briefed.

9.4.5 Procedure for case studies

To provide qualitative data concerning the process of learning while participants used *Link,* six students were initially assigned to act as case studies. This random assignment was made prior to the recruitment of participants. However, for one of the case studies a complete audio record was not made and so an additional participant was assigned to act as a case study prior to their recruitment. For a case study, the procedure for the *Link* group was followed as outlined above, but participants were also given the *Additional instructions* for the session which asked students to think aloud while they worked through the program. Participants were observed and a written and audio record of participants' comments while they used the program was made.

9.5 Findings

The average times to complete a pre test and a post test was 15 minutes and 10 minutes respectively. The average time it took students to work through *Link* was 17 minutes and participants took an average of 16 minutes to work through *Secrets of Psychology - Associative Learning*. Participants took an average of 15 minutes to complete the paper-based instructional materials.

9.5.1 Students' interactions

The user logs that were created when a student used *Link* were examined to determine the students' interactions with the program. Participants had been asked to find out about both studies and to complete the six learner activities. Specifically, students were asked firstly, to find out about the Health events study and complete the three activities for this study and then secondly, to find out about the Infant engagement study and complete the associated activities. The user logs indicated that these tasks were completed by the students and were also completed in the specified sequence. However, one participant's log indicated that she did not complete activity 1 for the Infant engagement study (S5) and part of two user logs were lost because two participants quit the program before they had finished completing all of the tasks. This meant that their original user file was over-written by a second user file that was created when they had completed all of the set tasks and therefore quit for a second time (S2, S5). The variety of student responses to the learner activities are summarised below. The three activities for the Health events section will be considered first followed by the activities for the Infant engagement study.

Health events study

Activity 1 asked students to select possible interpretations of a correlation that was found to exist between levels of depression and anxiety in the Health events study (figure 9.4). Participants completed this activity as follows:

- None of the participants selected the first option only, 'that the participants' depression caused them to be anxious'.
- One participant selected the second option only, 'that the participants' anxiety caused them to be depressed' (S3).
- Three participants selected the third option only, 'that the correlation between the depression and anxiety is spurious' (S26, S31, S34).

- Four participants selected the fourth option only, 'that another variable or variables could be responsible for the correlation' (S2, S8, S28, S30).
- Five participants selected the first and second options (S9, S12, S16, S18, S19).
- Two participants selected the third and fourth options (S13, S24).
- One participant selected the first, second, third and fourth options (S21).

It is noteworthy that only one participant chose to select all four possible interpretations of a correlation, which is the appropriate answer to the activity.

On activity 2, a user is asked which correlation in the displayed table represents the target scatter plot. A user can select any of the correlations in turn and they are provided with feedback in the form of text and a scatter plot (figure 9.8). Participants completed this activity in a variety of ways. For example, one participant selected the correlation 0.60 and then moved on to activity 3 (S18) and another participant selected all six correlations in turn until, on their sixth choice, they selected the correlation -0.35. Table 9.2 summarises the responses to this activity.

Table 9.2 Participant responses to activity 2. (Health events study)

	-0.35 selected first	-0.35 by third selection (< 4 choices)	-0.35 by sixth selection (≥ 4 choices)	-0.35 not selected
Number of Participants [a]	2 (S19, S24)	6 (S3, S8, S13, S21, S26, S34)	5 (S9, S12, S16, S30, S31)	2 (S18, S28)

[a] Missing data for two participants (S2, S5)

For activity 3 participants select the correlations from the data table and arrange them from the strongest to weakest relationship (figure 9.10). Table 9.3 shows participant responses to this activity.

Figure 9.10 Health events study. Activity 3

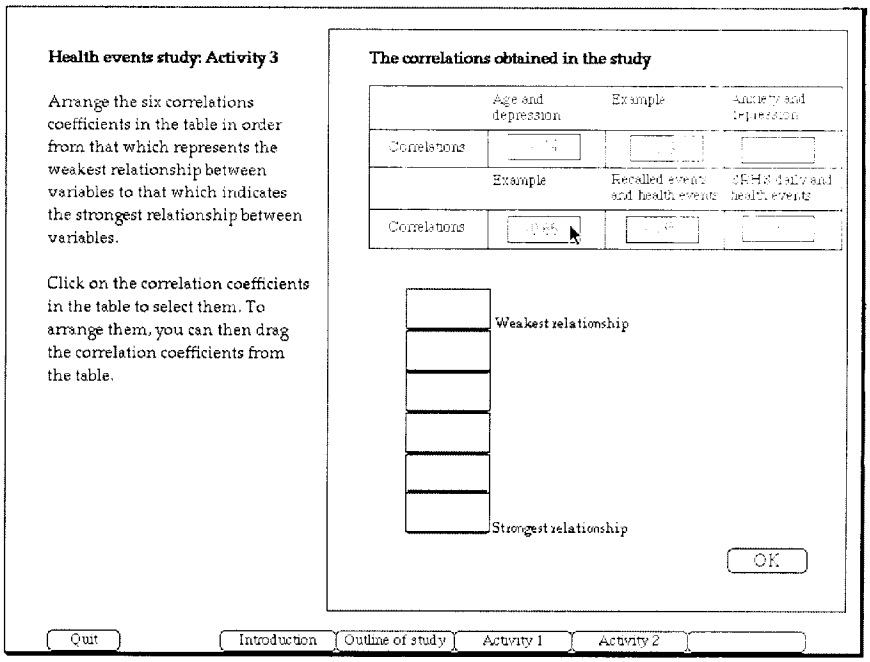


Table 9.3 Participant responses to activity 3. (Health events study)

	Correct arrangement	0.60 positioned as strongest	Other arrangements
Number of Participants [a]	10 (S2, S9, S13, S16, S19, S24, S16, S28, S30, S34)	4 (S12, S18, S31, S21)	2 (S3, S8)

[a] Missing data for one participant (S5)

It can be seen from table 9.3 that ten of the participants correctly arranged the correlation coefficients. The findings for the four participants who positioned 0.60 as the strongest relationship were as follows.

- Three participants (S12, S18, S31) arranged the correlations

  -0.65 weakest relationship
  -0.35
  -0.14
  0.18
  0.30
  0.60 Strongest relationship

- One participant (S21) arranged the correlations

  -0.14 weakest relationship
  -0.35
  -0.65
  0.18
  0.30
  0.60 Strongest relationship

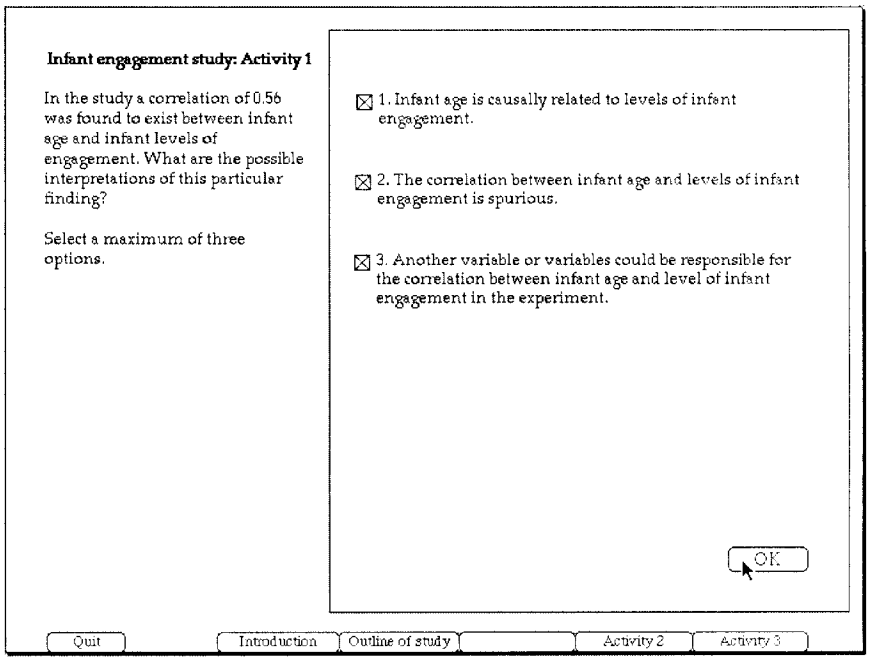
Infant engagement study

On activity 1 a user is asked what the possible interpretations of a correlation are that was found to exist between infant age and infant levels of engagement, and they are to select from three available options (figure 9.11). Participants completed this activity in the following ways:

- One participant selected the third option only, that 'Another variable or variables could be responsible for the correlation between infant age and infant engagement in the experiment' (S3).
- One participant selected the first and second options that 'Infant age is causally related to levels of infant engagement' and 'The correlation between infant age and levels of infant engagement is spurious' (S12).
- Five participants selected the first and third options (S2, S8, S16, S28, S31).
- One participant selected the second and third options (S13).
- Eight participants selected the first, second and third options (S9, S18, S19, S21, S24, S26, S30, S34).

It is noteworthy that eight out of the seventeen participants appropriately selected all three options or all possible interpretations of a correlation. On this activity for the Health events study, which was completed by participants before they completed activity 1 for the Infant engagement study, only one participant selected all four interpretations of a correlation.

Figure 9.11 Infant engagement study. Activity 1



In activity 2 for the Infant engagement study a user is required to select from six correlations the coefficient (-0.36) that represents the target scatter plot displayed on the screen. The findings relating to participant responses to this activity are shown in table 9.4, which shows that seven out of the seventeen participants first selected the correlation that represents the target scatter plot. This is in contrast to the findings relating to participant responses to activity 2 for the Health events study, which was completed by participants prior to them working on activity 2 for the Infant engagement study. When participants completed activity 2 for the Health events study, only two participants chose the target correlation -0.35 as the first selection.

Table 9.4 Participant responses to activity 2. (Infant engagement study)

	-0.36 selected first	-0.36 by third selection (< 4 choices)	-0.36 by seventh selection (≥ 4 choices)	-0.35 not selected
Number of Participants	7 (S3, S5, S12, S13, S16, S21, S24)	5 (S9, S19, S28, S30, S31)	4 (S2, S8, S26, S34)	1 (S18)

On activity 3 for the Infant engagement study, none of the seventeen participants placed the positive correlation 0.35 as the strongest relationship, rather than the negative correlation, -0.36 (figure 9.12). Table 9.5 summarises the participant responses to this activity.

Figure 9.12 Infant engagement study. Activity 3

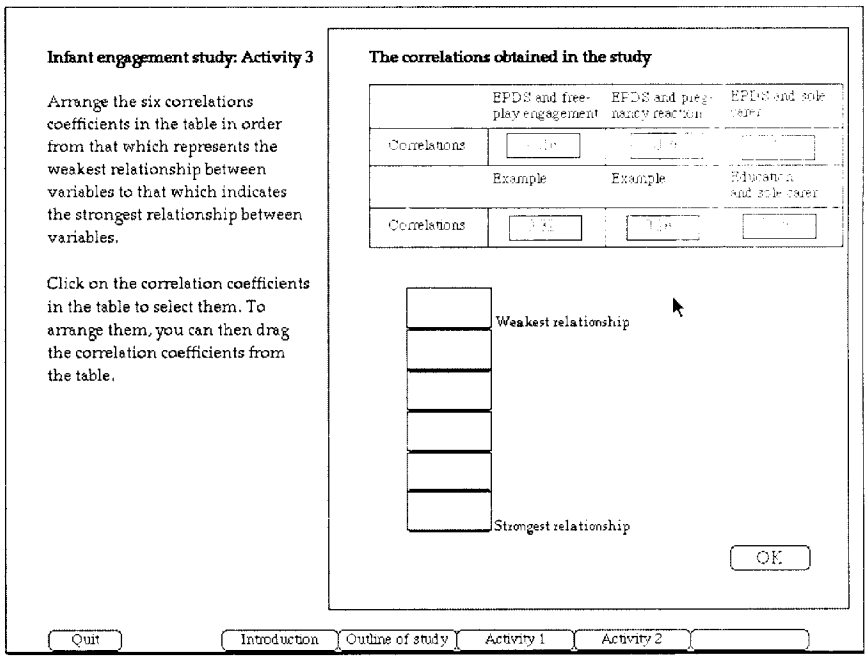


Table 9.5 Participant responses to activity 3. (Infant engagement study)

	Correct arrangement	0.32 positioned as strongest	Other arrangements
Number of Participants [a]	14 (S2, S5, S9, S12, S13, S16, S18, S19, S24, S26, S28, S30, S31, S34)	0	2 (S3, S8)

[a] Missing data for one participant (S21)

From table 9.5 it can be seen that fourteen out of the seventeen participants arranged all the correlations correctly from the weakest to strongest relationship. Participants completed activity 3 for the Health events study prior to this equivalent activity. It is noteworthy, therefore, that on activity 3 for the Health events study, ten participants

provided the correct arrangements of correlations and four participants positioned a positive correlation as stronger than a negative one.

9.5.2 Outcomes

The overall score on the pre and post tests provided an assessment of a participant's general understanding of correlation. Descriptive statistics for the pre test and post test scores for the three different groups are shown in table 9.6.

A sample of 24 pre and post tests (eight tests from the *Link* group, eight from the basic control group, and eight from the instructional control group) were scored by a second researcher at the Open University. An inter-rater reliability of 0.98 was obtained between the overall scores assigned by the thesis author and the scores assigned by the second researcher (d.f. = 22, $p < 0.001$).

As described above, seven of the participants in the *Link* group acted as case studies. These participants were observed and asked to think aloud while they used *Link* and it was thought that this procedure might have an impact on students' performance on the post test compared to those participants who did not act as case studies. The following analysis was therefore carried out.

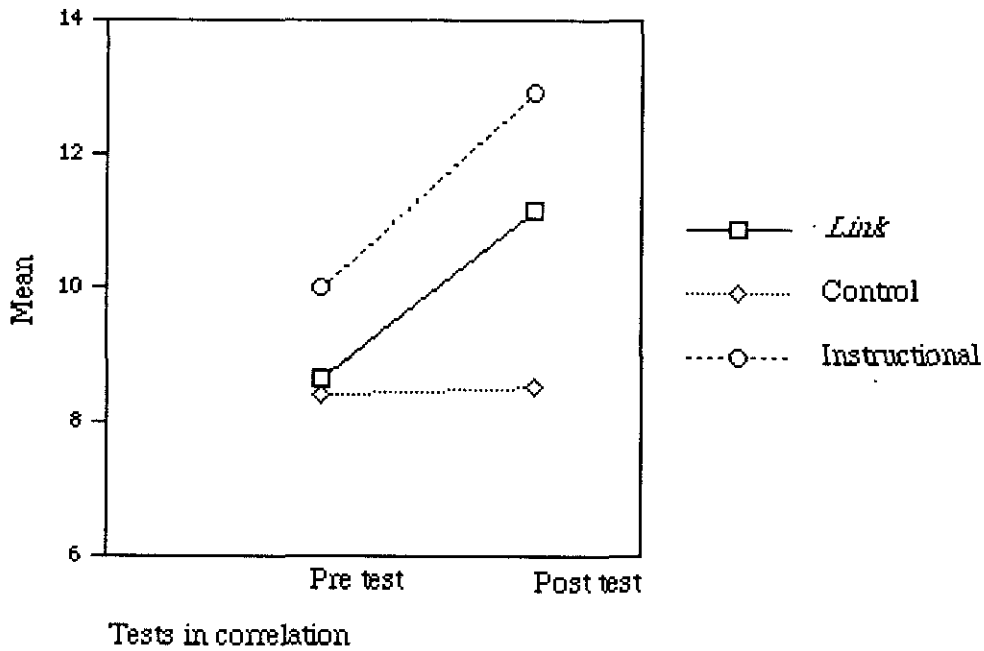
The pre test mean of those participants who acted as cases (mean = 9.57, S.D. = 5.34, n = 7) and the post test mean of this case study group (mean = 11.43, S.D. = 3.31, n = 7) were approximately equivalent to the pre test mean (mean = 8.00, S.D. = 4.45, n = 10) and post test mean (mean = 11.00, S.D. = 4.49, n = 10) of the *Link* group that excluded the case study participants. Case study participants were therefore included as cases in the subsequent statistical analysis described.

Table 9.6 Descriptive statistics for the pre and post tests

	Link group n = 17	Control group n = 17	Instructional control group n = 16
Pre test	mean 8.65 S.D. 4.73 min. 2.00 max. 17.00	mean 8.41 S.D. 4.05 min. 2.00 max. 13.00	mean 10.00 S.D. 5.39 min. 2.00 max. 17.00
Post test	mean 11.18 S.D. 3.94 min 0.00 max. 16.00	mean 8.53 S.D. 4.36 min 2.00 max. 16.00	mean 12.94 S.D. 3.39 min 3.00 max. 18.00

It can be seen from table 9.6 that the pre test mean of the instructional group is higher than that of the *Link* and control groups. However, the differences between the pre test means of the three groups were not significant ($F_{2, 47} = 0.53$, $p > 0.05$). The pre and post test means of the three groups is graphically presented in figure 9.13.

Figure 9.13 Pre and post test means of the three groups



An ANCOVA on post test scores, with pre test scores as a covariate, was carried out to see if there was a significance difference between the three groups. Table 9.7 provides the ANCOVA summary table and shows that the main effect of group was significant ($F_{2, 46} = 6.21$, $p < 0.01$). It can be concluded that after controlling for the individual differences in the pre test scores, the three groups do differ significantly on the post test.

Table 9.7 ANCOVA summary table

Source of variation	Degrees of freedom	Sum of squares	Mean square	F	Significance of F
Covariate Pre score	1	324.47	324.47	37.20	
Main effect Group	2	108.29	54.14	6.21	0.004
Error	46	401.18	8.72		
Model	3	487.54	162.51	18.63	
Total	49	888.72	18.14		

A contrast was carried out to see if there was a significant difference between the post test scores of the *Link* group and the instructional control group, but no significant difference was found ($t = 0.98$, $p > 0.05$.) A second contrast was conducted to see if the post tests scores of the *Link* group and the instructional control group, which were combined as groups for the contrast, differed significantly from the basic control group. In this case, a significant difference was found ($t = 3.41$, $p < 0.01$). With the exception of participants in the basic control group, there was an increase in students' scores from the pre test to the post test whether students used *Link* or the paper-based instructional materials. This is clear from figure 9.13 where the lines for the *Link* group and instructional control group are parallel.

It can therefore be concluded that compared to the basic control group, the participants' scores in both the *Link* group and the instructional control group significantly increased from the pre test to the post test. This suggests that the use of *Link* or the use of the paper-based instructional materials contributed to students' general understanding of correlation.

9.5.3 Students' misconceptions

The findings relating to students' misconceptions in the *Link* and instructional control group are reported here because it is these groups that showed a significant increase in mean scores from pre to post test.

Link

For the *Link* group, table 9.8 shows the different categories for students' answers to questions on the pre test. This table presents the proportion of participants' answers that were coded for each category for seven out of the ten questions on the pre test. A question was included in this table if two or more participant responses to a question were categorised as a particular misconception.

Table 9.8 Participant responses to questions on the pre test (Link group)

	Qu 1	Qu 2	Qu 4	Qu 5	Qu 7	Qu 8	Qu 10
Appropriate answer	64.71% (11) [e]	58.82% (10)	47.06% (8)	5.88% (1)	29.41% (5)	17.65% (3)	5.88% (1)
Causalistic	11.76% (2)	N/A	11.76% (2)	N/A	N/A	N/A	N/A
Unidirectional	N/A	29.41% (5)	N/A	11.76% (2)	11.76% (2)	N/A	N/A
Strength [a]	N/A	N/A	N/A	N/A	5.88% (1)	23.53% (4)	41.18% (7)
Lacks prior knowledge [b]	17.65% (3)	(0)	(0)	11.76% (2)	(0)	5.88% (1)	5.88% (1)
Explanation insufficient [c]	(0)	(0)	5.88% (1)	23.53% (4)	23.53% (4)	35.29% (6)	35.29% (6)
Idiosyncratic [d]	5.88% (1)	11.76% (2)	35.29% (6)	47.06% (8)	29.41% (5)	17.65% (3)	11.76% (2)

[a] The conception that a positive correlation is stronger than a negative correlation
[b] If a student left the answer blank or wrote "don't know".
[c] The student's explanation to the question was insufficient.
[d] If a student provided an idiosyncratic response e.g., a correlation coefficient is interpreted as a level of probability.
[e] Frequencies are given in brackets.

Question 1 and question 4 on the tests were designed to test a student's understanding of correlation and causation. It can be seen from table 9.8 that for both these questions 11.76 per cent of the participants' answers were categorised as a causalistic conception.

Question 2, question 5 and question 7 were designed to identify a unidirectional conception of correlation. On question 2, five of the participants' answers were categorised as a unidirectional conception of correlation. Question 2 presented a scatter plot and asked what the scatter plot shows about the relationship between two sets of scores. Although over half (58.82%) of the participants in the *Link* group correctly described the relationship displayed by the scatter plot as a negative correlation, out of the five participants whose answers to this question were coded as a unidirectional

conception of correlation, three of the participants described a positive correlation (S2, S8) and two of the participants simply indicated that a relationship existed, but they did not specify the direction (S12, S16, S18). For example, one participant answered question 2 as follows:

> "Children with higher arithmetic scores have higher reading scores" (S8).

And one participant answered the question by writing:

> "It shows that memory + spelling scores can be said to be of a high correlation and that memory may affect spelling" (S12).

Only two of the participants' answers to question 5 were coded as a unidirectional conception of correlation (S21, S31). Question 5 asked what value of a correlation coefficient indicates no relationship between two variables and one of these two participants wrote:

> "A (-) value indicates a negative (inverse) relationship" (S31).

However, this participant's answer to question 2 was categorised as an appropriate answer because he had described a negative relationship in which

> "The higher the score on the memory test the lower the scores on the spelling test (negative correlation)" (S31).

The other participant (S21) whose answers to question 5 was coded as a unidirectional conception gave an equivalent pattern of response: to question 5 she provided the value -0.02 as indicating no relationship, and explained her answer by stating that

> "This number is far away from 1, and a negative value" (S21).

However, in her answer to question 2, she correctly wrote that there was a negative relationship between spelling and memory test scores. This pattern of responses to question 2 and question 5 could either indicate that a student's unidirectional conception is not necessarily stable or that a student holds ideas that are contradictory.

A majority of the participants' answers to question 5 were categorised as idiosyncratic. For example, on question 5, which asks for a value of a correlation that would indicate no relationship between variables, one participant wrote:

> "More that 0.5"
> Explain your answer
> "Leaves a big percentage that could show us that the results are due to chance" (S3).

Two of the participants' answers to question 7 were categorised as a unidirectional conception (S18, S31).

Question 8 and question 10 were designed to identify the conception that a positive correlation is stronger than a negative correlation. This question asked which of two correlations is stronger (e.g., 0.73 or -0.84) and four out of the seventeen participants' answers were coded as the strength misconception because they indicated that the positive correlation was stronger than the negative one (S5, S12, S18, S30). One of these participants explained his answer by writing:

> "Because it shows a positive correlation" (S30).

Question 10 asked participants which set out of four sets of correlation coefficients correctly shows the strongest to weakest relationship. On this question, forty one percent of the participants' answers were categorised as the strength misconception (S5, S8, S12, S18, S21, S30, S31). Six out of these seven participants (S5, S12, S18, S21, S30, S31) indicated the set in which the (strong) negative correlation was the weakest relationship as follows:

A majority of the participants' answers to question 5 were categorised as idiosyncratic. For example, on question 5, which asks for a value of a correlation that would indicate no relationship between variables, one participant wrote:

> "More that 0.5"
> Explain your answer
> "Leaves a big percentage that could show us that the results are due to chance" (S3).

Two of the participants' answers to question 7 were categorised as a unidirectional conception (S18, S31).

Question 8 and question 10 were designed to identify the conception that a positive correlation is stronger than a negative correlation. This question asked which of two correlations is stronger (e.g., 0.73 or -0.84) and four out of the seventeen participants' answers were coded as the strength misconception because they indicated that the positive correlation was stronger than the negative one (S5, S12, S18, S30). One of these participants explained his answer by writing:

> "Because it shows a positive correlation" (S30).

Question 10 asked participants which set out of four sets of correlation coefficients correctly shows the strongest to weakest relationship. On this question, forty one percent of the participants' answers were categorised as the strength misconception (S5, S8, S12, S18, S21, S30, S31). Six out of these seven participants (S5, S12, S18, S21, S30, S31) indicated the set in which the (strong) negative correlation was the weakest relationship as follows:

Strongest weakest
0.83, 0.65, 0.03, -0.91

Explanations to this choice ran as follows:

"Correlation coefficient becomes more negative therefore weakest" (S5).

"Because it ranges from negative to positive" (S21).

"Because the weakest always starts at a negative to a positive" (S30).

One of the seven participants (S8) indicated the set in which no correlation (e.g., 0.04) is the weakest relationship, but that the (strong) negative correlation is weaker than the positive correlation:

Strongest weakest
0.67, 0.56, -0.79, 0.04

From table 9.8 it can be seen that for questions 5, 7, 8, and 10 more than twenty per cent of the participants' responses could not be categorised as normal or as a particular misconception because their explanations to the questions were not sufficient to determine if they held a particular conception. This had implications for the McNemar analysis described below: it meant that those participants' answers that were not categorised as normal or as a misconception could not be included as cases in the McNemar tests for change for the questions. Similarly, table 9.8 shows that for questions 4, 5 and 7 more than twenty per cent of the participants' responses were categorised as idiosyncratic and could not therefore be included as cases in the McNemar tests for change.

The category idiosyncratic includes responses that are described as additional confusions held by participants, which are considered below, but also includes other responses that were not possible to code in any other way. For example, question 2 asked what a scatter plot shows about the relationship between two sets of scores, and the appropriate answer to this question is to state that the scatter plot indicates a negative correlation. However, in response to this question, a participant's answer was

categorised as idiosyncratic because it read as follows:

"The distribution of scores and whether there is an interaction" (S28).

From an examination of students' answers on the pre test, three additional kinds of confusions or misconceptions were evident. Firstly, one participant (S2) thought that correlations near 1 indicated no relationship between variables and that a correlation near zero indicated a strong correlation. This participant answered questions 5, 6, 7, 9 and 10 accordingly. For example, question 6 asked which correlation, -0.88 or 0.02 is stronger and to this she indicated that 0.02 was the stronger correlation and explained:

"Further away from 1.0" (S2).

Secondly, one participant (S31) held a confusion relating to variance and his answers to question 3 and question 6 were coded as idiosyncratic. On question 6, this participant indicated that the correlation 0.02 was stronger than -0.88 and explained:

"Because the level of variance or co-variation is positive (converse)" (S31).

Thirdly, five out of the seventeen participants in the *Link* group were confused by significance levels (S3, S18, S12, S18, S24). It is noteworthy that none of the questions on the tests included the words significant or probability. It was evident that participants had such a confusion because their answers to questions on the pre test referred to the significance levels of 0.01 and 0.05 that are typically used in psychological research. For example, on question 9 one participant (S3) decided that 0.02 indicated a relationship between variables (rather than -0.82) and wrote:

"It is between 0.01 and 0.05 so it is acceptable" (S3).

Similarly, on test D, on question 9 one participant (S24) decided that 0.01 indicated a relationship between variables and wrote that this

"shows that the relationship between the 2 variables is highly significant" (S24).

When asked on question 3 which correlation coefficient, 0.03 or 0.68 was stronger, one participant (S8) indicated 0.03 and explained that

"0.68 is greater than 0.5. 0.03 is less than 0.5 making it a significant value" (S8).

Participants' answers to the questions on the post test were also categorised to identify appropriate answers or misconceptions. One-tailed McNemar tests were carried out on questions 1, 2, and 4 to see if the frequencies of students' answers that were categorised as causalistic or unidirectional on the pre test changed on the post test. These tests indicated that there were no significant changes in the frequencies of particular responses from the pre test to the post test (question 1, cases 13, Binomial = 0.95; question 2, cases 14, Binomial = 0.5; question 4, cases 9, Binomial = 0.5). McNemar tests are not reported for questions 3, 5, 6, 7, 8, 9 and 10 because the number of cases that could be included in the analysis for each question was less than six.

These findings indicate that having used *Link*, participants' misconceptions did not necessarily change. Although there were no significant changes in the frequencies of participants' answers that had been coded as a particular misconception (e.g., unidirectional) from the pre test to the post test, there were cases in which participants' misconceptions were addressed by the learner activities. This is clear from the participants' answers to questions on the pre and post tests, and how they completed the learner activities.

One participant's (S31) answer to question 4 on the pre test was categorised as a causalistic conception, but on the post test this participant's answer was coded as appropriate where he explained:

> "Cannot claim causal relationship - there may be extraneous variables affecting the findings" (S31).

Examination of participants' answers on the pre test revealed that five of the participants held a unidirectional conception of correlation (S8, S12, S18, S21, S31). The participants' answers on the post test showed that out of these five participants, only two participants still held a unidirectional conception (S8, S12) and three of the participants no

longer held this conception (S18, S21, S31). For example, on the pre test, a participant answered question 5 by writing that a negative value would indicate no relationship between two variables and explained that it indicated a negative relationship (S31). However, on the post test this participant answered question 5 by answering the question correctly as follows:

"0.00. A value close to 0 indicates a weak relationship" (S31).

On question 10(i) on the pre test, seven of the seventeen participants indicated that a positive correlation is stronger than a negative correlation (S5, S8, S12, S18, S21, S30, S31). Of these seven participants, four participants arranged the set of correlations on activity 3 for the Health event study so that the 0.60 was incorrectly placed as the strongest relationship rather than -0.65 (S12, S18, S21, S31), one participant incorrectly placed 0.18 as the strongest relationship (S8), one participant provided a correct arrangement of correlations (S30), and data from one participant was not recorded (S5). On activity 3 for the Infant engagement study, only one of the seven participants arranged the correlations incorrectly (S8). Data for this activity was not recorded for one participant (S21). This meant that five of the seven participants arranged the correlations correctly (S5, S12, S18, S30, S31).

Having used *Link*, four of the seven participants who on the pre test indicated that a positive correlation was stronger than a negative correlation, correctly indicated the set of correlations that showed the weakest to strongest relationship on the post test (S5, S18, S21, S30).

Three additional confusions held by participants in the *Link* group, which were indicated by their answers to questions on the pre test, were not necessarily evident from their answers on the post test. It was outlined above that one participant (S2) thought that correlations near 1 indicated no relationship between variables and that a correlation near zero indicated a strong correlation. On the post test, this student no longer expressed this confusion, where on question 6, for example, she indicated that the correlation -0.82 was

stronger than 0.04 and explained that it was a higher number nearer one. It was also noted above that one participant (S31) held a confusion relating to variance, but this participant's answers to questions on the post test did not indicate this. Of the five participants who were confused by significance levels, three of these participants did not hold this confusion after using *Link*. This was indicated by their answers to questions on the post test (S12, S18, S24). However, having used *Link*, two of these five participants still were confused by significance levels (S3, S8).

In summary, the participants in the *Link* group held particular misconceptions that the program was designed to address, but also lacked the necessary prior knowledge to answer questions, or held additional confusions that, for example, related to levels of statistical significance. In addition, it was not always possible to determine if participants held an appropriate conception because they did not provide sufficient explanations to particular questions.

Although *Link* contributed to students' general understanding of correlation, it is not clear whether the program addressed and changed particular misconceptions. This is because participants did not necessarily hold the misconceptions that the program was designed to address (e.g., causalistic), but held additional conceptions that were categorised as idiosyncratic or lacked the necessary prior knowledge.

Instructional control

For the instructional control group, table 9.9 shows the different categories for students' answers to questions on the pre test. This table presents the proportion of participants' answers that were coded for each category for seven out of the ten questions on the pre test.

Table 9.9 Participant responses to questions on the pre test (instructional control group)

	Qu 1	Qu 2	Qu 4	Qu 5	Qu 7	Qu 8	Qu 10
Appropriate answer	56.25% (9) [e]	75% (12)	43.75% (7)	12.50% (2)	25% (4)	37.50% (6)	6.25% (1)
Causalistic	25% (4)	N/A	31.25% (5)	N/A	N/A	N/A	N/A
Unidirectional	N/A	6.25% (1)	N/A	(0)	(0)	N/A	N/A
Strength [a]	N/A	N/A	N/A	N/A	(0)	25% (4)	25% (4)
Lacks prior knowledge [b]	6.25% (1)	6.25% (1)	12.50% (2)	31.25% (5)	25% (4)	6.25% (1)	12.50% (2)
Explanation insufficient [c]	12.50 (2)	6.25% (1)	6.25% (1)	37.50% (6)	31.25% (5)	25% (4)	50% (8)
Idiosyncratic [d]	(0)	6.25% (1)	6.25% (1)	18.75% (3)	18.75% (3)	6.25% (1)	6.25% (1)

[a] The conception that a positive correlation is stronger than a negative correlation
[b] If a student left the answer blank or wrote "don't know".
[c] The student's explanation to the question was insufficient.
[d] If a student provided an idiosyncratic response e.g., a correlation coefficient is interpreted as a level of probability.
[e] Frequencies are given in brackets.

Compared to the *Link* group, the pattern of participants' misconceptions in the instructional control group is different.

Question 1 and question 4 on the tests were designed to identify a causalistic conception of correlation. Table 9.9 shows that on question 1 on the pre test twenty five per cent of the participants' answers were categorised as causalistic (S38, S39, S45, S47) and on question 4, thirty one per cent were categorised in this way (S35, S38, S41, S44, S45). For example, question 4 gives a correlation of 0.87 between length of time a person is in prison and amount of aggression displayed on a psychological inventory and asked:

This means that spending a longer amount of time in prison causes people to become more aggressive. True or false?

In response to this question, one participant indicated 'true' and wrote:

"Being locked up in a confined space for long periods of time is likely to lead to increased aggression, due to frustration, and acclimatisation to a male dominated, female free, aggression based hierarchical system" (S44).

Question 2, question 5 and question 7 were designed to identify a unidirectional conception of correlation. Table 9.9 shows that in contrast to the *Link* group, only one of the participant's answers was categorised as a unidirectional conception on the pre test (S39). In her answer to question 2, in which data on a scatter plot should be interpreted as a negative correlation, this student wrote:

"The scatter plot shows that there is no real relationship between the reading and arithmetic scores. This is because there are children with reading scores of 1 with an arithmetic score of 14. Therefore is it down to individual ability" (S39).

As is considered below, this participant also held confusions relating to statistical significance.

Question 8 and question 10 were used to test a participant's understanding of the strength of correlations and were used therefore to identify the strength misconception. On question 8, a quarter of the participants' answers were categorised as this misconception (S35, S38, S39, S41) and a quarter of the participants' answers to question 10 were coded as this misconception (S35, S38, S39, S41). For question 10, all of these four participants selected the set of correlations in which a (strong) negative correlation represented the weakest relationship (e.g., 0.83, 0.65, 0.03, -0.91).

With regard to participants' answers on the pre test, only two participants in the instructional control group held confusions that were not to do with the misconceptions described above (e.g., causalistic). One participant thought that correlations near 1

indicated no relationship (S46) and in contrast to the *Link* group, only one participant was confused by levels of statistical significance (S39). In answering question 3, for example, this participant indicated that 0.01 was stronger than 0.64 and explained:

> "With a correlation co-efficient of 0.01 there is only a 1% chance of getting the results wrong" (S39).

Participants' answers to the questions on the post test were also categorised to identify appropriate answers or misconceptions. One-tailed McNemar tests were carried out on questions 1 and 4 to see if the frequencies of students' answers that were categorised as causalistic on the pre test changed on the post test. These tests indicated that there were no significant changes in the frequencies of particular responses from the pre test to the post test (question 1, cases 10, Binomial = 0.50; question 4, cases 11, Binomial = 0.25). McNemar tests are not reported for questions 2, 3, 5, 6, 7, 8, 9 and 10 because the number of cases that could be included in the analysis for each question was less than six. These findings indicate that having completed paper-based instructional materials, participants' misconceptions did not necessarily change.

Additional confusions held by the two participants in the instructional control group, which were indicated by answers on the pre test, were not evident from their answers to questions on the post test (S39, S46). However, in contrast to one participant's answers on the pre test, a participant's answers to two of the questions on the post test indicated that he was confused by the term 'coefficient'. On question 6, for example, this participant indicated that the correlation coefficient 0.04 was stronger than -0.82 and explained:

> "Correlation coefficient means less correlation. Closer to 0" (S37).

In summary, the profile of participants' misconceptions in the instructional control group was different to the participants in the *Link* group. For example, in the former group only one of the participant's answers to question 2 was categorised as unidirectional, whereas in the *Link* group five of the participants' responses were

categorised in this way.

9.5.4 Case studies

In the *Link* group, seven of the participants were randomly assigned to act as case studies (S2, S3, S13, S16, S19, S24, S26). To illustrate how the learner activities in *Link* affected students' understanding of correlation three of these cases are considered. These three case studies were selected to illustrate how participants differed with regard to their pre and post test score and their completion of the learner activities: two of the case studies (S2, S3) obtained a relatively low pre test score, but obtained a higher post test score, and one of the case studies (S16) obtained the same score on both the pre and post test. Participants names have been changed to assure anonymity.

Olive (S2) obtained a score of 4 out of 19 on the pre test and her answer to question 2 on this test was coded as a unidirectional conception. This question was answered as follows:

"The better the memory test results the better the spelling test results" (S2).

The pre test indicated that Olive thought that correlations near 1 indicated no relationship between variables and that a correlation near zero indicated a strong correlation and she answered questions 5, 6, 7, 9 and 10 accordingly. For example, question 5 asked what value of a correlation coefficient indicates no relationship between variables, and Olive answered this by writing "0.9" and explained "near to 1.0". On question 7, Olive indicated that neither 0.68 or -0.85 showed a relationship between variables and explained that they were "near to 1.0". Similarly, on question 9, she selected 0.01 (rather than -0.86) as indicating a relationship and wrote "further away from 1.0". On question 10 she chose the following set of correlations as representing the weakest to strongest relationship:

Weakest strongest
-0.79, 0.56, 0.67, 0.04

In using *Link*, Olive found out about the Health events study by reading the text on the screen that provided details of this study. On activity 1, Olive selected the fourth option 'that another variable or variables could be responsible for the correlation', but in response to the audio feedback she asked "Why are you repeating things to me?" One part of the feedback to activity 1 stated that 'It is possible that the participants' anxiety caused them to be depressed' and in response to this Olive commented:

> "Yes it is possible. But it doesn't say it is possible at the beginning. Why are you saying it is possible ... It is possible, but that isn't the question, is it?"

The question for activity 1 does ask "What are the possible interpretations of this particular finding?"

Having read the question for activity 2 and having looked at the correlations in the table, Olive remarked "I wonder what those minuses mean". On activity 2, students select the correlation in the table that represents the data on the displayed scatter plot. With regard to the target scatter plot, Olive commented:

> "I would have said ... It's all over the shop. So there isn't anything. The dots are all over the place. There's not a nice little line of dots anywhere".

Olive then went on to say that she was "just kind of guessing now" and reiterated "God, what do these minuses mean?" and she then selected the correlation -0.30 and the scatter plot that represents this relationship was displayed on the screen along side the target scatter plot. Olive read the text-based feedback associated with this selection and commented:

> "Oh right. So a positive means they're both going the same way and a negative means one's different to the other. I still don't really understand it".

Olive then selected the correlations 0.60, 0.30 and -0.14 in turn and went on to say

> "So the bigger the number the better the correlation whether it's a minus or plus. And minuses mean it's a negative correlation so both go in different directions ... I'm getting there now".

In response to reading the question for activity 3 Olive said

> "I should know this now we've just learnt it ... But what about the negative and the positive. Does it matter or not?"

Olive then successfully arranged the correlations in the correct order and while doing this spoke out aloud:

> "It was the biggest number is the strongest thingy wasn't it. So the littlest number is ... the weakest ... so it doesn't matter if it's a minus or a plus".

As instructed, Olive then read about the Infant engagement study and completed the three activities for this study. On activity 1 she selected all of the available options, but did ask, "What's causally related?" She also commented, "It's not the cause, but it could be". On activity 2 Olive initially selected the correlation 0.50 as representing the target scatter plot, then said, "Oh that's not right". She went on to select -0.16 and 0.11 and then she decided "Oh give up ... I don't know what it is". She did, however, go on to select 0.60, 0.56 and then the target correlation, -0.36. On activity 3 for the Infant engagement study, Olive again arranged the set of correlations in the correct order.

Olive obtained a score of 10 out of 19 on the post test. Her answers on this test showed that she no longer thought that a correlation near 1 indicated no relationship and that a correlation near zero indicated a strong correlation. However, as on the pre test, her answer to question 2 was categorised as a unidirectional conception in which she wrote "strong positive correlation" under the scatter plot for question 2 which showed a negative relationship. In spite of Olive correctly arranging sets of correlations from the strongest to weakest relationship when she used *Link*, Olive answered question 10 on the

post test as follows:

Strongest weakest
0.83, 0.65, -0.91, 0.03

"Highest to lowest numbers".

The second case study, Carol, obtained a score of 2 on the pre test out of a possible 19 (S3). Here, her answer to question 1 was coded as causalistic and her answers to questions 3, 5, 6, 7 and 9 indicated that she held a confusion relating to levels of statistical significance. For example, on question 3 Carol indicated that the correlation 0.01 was stronger than 0.64 and explained:

> "Because it leaves just 1 chance out of a 100 that our results might not be right or that it might be due to chance".

When using *Link*, Carol read the screen that provides an outline of the Health events study, and then went on to complete activity 1. Having read the question for this activity Carol commented, "take a wild guess that's what I'll do", and she selected the second option, 'that the participants' anxiety caused them to be depressed'. Carol read and listened to the feedback to this activity, but commented:

> "Oh god I made a mistake ... How was I supposed to know ... I'm not good at this anyway".

Having read what she had to do for activity 2, Carol went on to comment:

> "What? ... I have no idea ... I don't know ... I think it might be a minus, but I'm not sure".

Carol then selected the correlation -0.21 and when the associated scatter plot was displayed on the screen alongside the target scatter plot, she remarked "it's not", but she then selected the correlation that represented the target scatter plot and remarked:

> "Oh I got it ... I knew it was a negative one, but I didn't know which one".

On activity 3, Carol arranged the set of correlations as follows:

-0.65 Weakest relationship
-0.35
-0.14
0.60
0.30
0.18 Strongest relationship

When Carol had received feedback to this arrangement, she went on to say:

> "Are these the results? Oh god I would have failed my R and E classes [Research and Experimentation course] ... Not a single one is right ... So I suppose minus or sort of zero doesn't have anything to do with how weak or strong a correlation is".

Carol went on to read about the Infant engagement study and completed activity 1 for this study by selecting all of the available three options. When Carol completed activity 2 for the Infant engagement study she remarked, "so this one's the same as before", and selected the correlation, -0.36 that represented the target scatter plot on her first and only choice. On beginning to complete activity 3, Carol commented, "this one I messed up before", and she arranged the correlations incorrectly as follows:

-0.36 Weakest relationship
0.32
0.26
-0.16
0.11
0.05 Strongest relationship

Carol obtained a score of 7 out of 19 on the post test, but her answer to question 4 was categorised as causalistic and her answers to questions 3, 5, 6, 9 showed that she was confused. For example, on question 3 she indicated that the correlation 0.03 was stronger than 0.68 and explained her answer as follows:

> "I don't know. I'm confused from what I read on the computer of course this was about relationship but I can't tell for sure".

And in response to question 5 that asked for a correlation coefficient which indicates no relationship, Carol wrote:

> "I don't know. It seems that I've learned something wrongly and it confused me a lot".

However, on question 8, Carol correctly indicated that -0.84 was stronger than 0.73, and explained

> "0.84 > 0.73. Minus is not taken into consideration".

The third case study, Simon, obtained a score of 13 out of 19 on his pre test (S16), and his answer to question 2 on this test was categorised as unidirectional.

As instructed when using *Link*, Simon read about the Health events study and then went on to complete activity 1 for this study. Prior to selecting the first option on activity 1, 'that the participants' depression caused them to be anxious', and the second option, 'that the participants' anxiety caused them to be depressed', Simon commented, "just thinking about the correlation ... think it's quite strong". The correlation on this activity is 0.60.

Having read the question for activity 2 where a student has to decide which correlation coefficient represents a scatter plot, Simon said "so just ... looking for a weak correlation ... Think that's the weakest one" and he selected -0.14. When Simon had viewed the scatter plot associated with this correlation he commented:

> "So I think it's got to be a positive value but a small one as well".

Simon then went on to select 0.28, 0.30, 0.60, -0.21 and finally, -0.35 in turn.

On completing activity 3, Simon said that he thought the correlation -0.14 was the weakest correlation and he proceeded to arrange the correlations in the correct order.

Simon read about the Infant engagement study and went on to complete activity 1. On this activity he commented that the coefficient, 0.56 was "not particularly strong". He

selected option 1, that 'Infant age is causally related to levels of infant engagement', and the third option, that 'Another variable or variables could be responsible for the correlation between infant age and infant engagement in the experiment'.

On activity 2 for the Infant engagement study, Simon initially selected the correlation -0.36 that represented the target scatter plot and prior to this selection commented:

> "Again I don't think the correlation doesn't seem to be that strong, but they're all in sort of in the same area so it can't be the weakest one either ... I think it's a negative correlation".

On activity 3, Simon correctly arranged the correlations from the strongest to weakest relationship.

Simon obtained the same score of 13 on his post test as he obtained on his pre test. In contrast to his answer to question 2 on the pre test, which was coded as unidirectional, his answer to this question on the post test was exemplary:

> "The graph shows there to be a negative correlation, which is quite strong, because the plots are close to forming a straight line".

However, unlike his answer to question 4 on the pre test his answer to this question on the post test was categorised as causalistic. Simon did not obtain correct scores for particular questions because he did not always provide sufficient explanations. For example, on question 10 on the post test Simon correctly indicated the set of correlations that showed the strongest to weakest relationship, but his explanation to this simply read:

> "-0.91 is the closest figure to -1, and 0.03 the closest to zero".

A complete explanation to question 10 needs also to refer to 1 and therefore explain why, say, -0.91 represents the strongest relationship, but a correlation of 0.83 does not.

The three case studies illustrate that participants did hold particular misconceptions about correlation that *Link* was designed to address. However, participants also held additional confusions that the program did not cover. For example, Olive thought that correlations near 1 indicated no relationship and Carol held a confusion that related to levels of significance. Olive's answers on the post test showed that she no longer thought that a correlation of near 1 indicated no relationship, but her answer to question 2 was categorised as unidirectional as it was on the pre test. Carol's answers to questions on the post test indicated that she was still generally confused. However, although Simon's answer to question 2 on the pre test was categorised as unidirectional, his answer to this question on the post test was appropriate.

9.6 Discussion

The summative evaluation study provided data concerning students' general understanding of correlation and was designed to investigate whether the use of *Link* contributed to this understanding. Findings showed that with respect to the *Link* group, participants' scores on the tests in correlation increased from the pre test to the post test. However, findings also indicated there was an increase in the pre to post test scores for those students in the instructional control group. Concurrently, the participants' scores in the basic control group did not increase from the pre to the post test. An ANCOVA on post test scores, with pre tests scores as the covariate, was carried out and the main effect of group was found to be significant, meaning that there was a significant difference between the three groups. Statistical contrasts were undertaken and it was found that there was no significant difference between the post test scores of the *Link* group and the instructional control group, but that when these two groups were combined for analysis, it was found that the *Link* and instructional control group post tests scores differed significantly from the basic control group. These findings suggest that the use of *Link* contributed to students' general understanding of correlation, but that the use of the paper-based instructional materials, which were devised for the study, could also achieve

this.

Link was specifically designed to address particular misconceptions about correlation that had been identified through previous research (Morris, 1997). In *Link*, three kinds of learner activities are provided to address these misconceptions: a causalistic conception, a unidirectional conception and the conception that a positive correlation is stronger than a negative correlation. It was therefore thought that by using *Link* students' misconceptions would change because they would complete learner activities which provided relevant feedback to the student. However, in the summative evaluation it was found that participants in both the *Link* group and instructional control group not only held misconceptions that the questions on the pre and post test were designed to identify, but also lacked the necessary prior knowledge to answer questions, or held additional confusions that, for example, related to levels of statistical significance. As this was the case, it is suggested that introductory learning materials on the topic of correlation would have contributed to students' general understanding of correlation, whether these materials are a computer-assisted learning program or paper-based instructional materials. The important issue here is that *Link* was designed to address specific misconceptions about correlation, but participants who took part in the study did not necessarily hold these misconceptions. Instead they were confused by other matters, such as levels of statistical significance. The participants' general understanding of correlation would have improved if they worked through text-based or computer-based instructional materials that covered correlation. The implication of this is that the *use* of a computer-assisted learning program, which is specifically designed to target misconceptions in a topic area, must be carefully considered. In the case of *Link*, pre tests could be used to identify students' misconceptions and the program could be appropriately used for instructional purposes if the students were found to hold any of the misconceptions that the program was designed to address.

It should be emphasised that even though it took students only an average of just under twenty minutes to work through the activities in *Link*, participants' general understanding of correlation was improved. The above finding that participants' general understanding of correlation was improved, whether they used *Link* or completed instructional materials, are comparable to findings of evaluation studies of *Stat Lady*, which were reviewed in chapter 4 (e.g., Shute & Gawlick-Grendell, 1994). Shute and Gawlick-Grendell (1994) found that in terms of learning outcomes participants who used *Stat Lady* performed as well as those students who completed a paper-and-pencil Workbook version of the curriculum covered by *Stat Lady*. Specifically, performance on the learning outcome measure did not differ significantly between the *Stat Lady* group and the Workbook group, but both of these groups performed significantly better than the control group (Shute & Gawlick-Grendell, 1994). Shute and Gawlick-Grendell (1994) were encouraged by such findings because they argued that people from the general population are less familiar with computer technology and have experience in learning from textbooks or workbooks.

Computer-assisted learning programs were not used by the students, who participated in the study at Luton University, on the Research and Experimentation courses that they completed at the university to learn about statistics in psychology. Although the students at Luton were not familiar with learning statistics from a computer-assisted learning program, the findings of the summative study suggest that the use of *Link* did successfully contribute to students' general understanding of correlation.

The summative evaluation provided comprehensive qualitative data concerning students' misconceptions. This data extends the findings of the studies that were reported in chapter 3 and chapter 8. The findings of the summative study indicated that psychology students are confused by negative correlations, the strength of correlations and causality, but that students also hold additional misconceptions. The findings relating to students' misconceptions also indicated that particular misconceptions were not always stable or

that students held ideas that were contradictory. For example, two students in the *Link* group thought that a negative correlation indicated no relationship, but they both successfully interpreted a scatter plot as showing a negative relationship.

Participants' answers on the pre and post tests were categorised as misconceptions or as appropriate and McNemar tests was carried out to see if there were significant changes in the frequencies of students' responses to particular questions from the pre test to the post test. For both the *Link* and instructional control groups, none of the these statistical tests were significant. However, in the *Link* group, for example, McNemar tests were not reported for seven out of the ten questions because less than six cases could be included in the analysis. This was because only a handful of cases were categorised as an appropriate answer or as a particular misconception (e.g., causalistic) for such questions, but otherwise participants' answers were coded as 'lacks prior knowledge' or as idiosyncratic, or the participant's answer could not be categorised as normal because their explanation was insufficient. In particular cases, the idiosyncratic category accounted for a confusion that related to levels of statistical significance. The tests in correlation did not include the words significance or probability, but on a question a participant would, for example, interpret that the correlation 0.01 was stronger than another correlation because it was significant.

The summative study was designed to investigate whether the learner activities in *Link* affected students' misconceptions relating to correlation. An analysis of the user logs indicated that there was a variety of student responses to the learner activities, which were summarised. From this it was clear that in general, participants' performance at completing the learner activities improved as they used *Link*. All of the participants had to firstly find out about the Health events study and complete the three activities for this study and then to find out about the Infant engagement study and complete the activities for this study. It is interesting to find that in the case of activity 1 for the Health events study, only one of the participants chose to select all four possible interpretations of a

correlation, but subsequently eight of the participants selected all of the possible interpretations of a correlation relating to the Infant engagement study. Similarly, the target scatter plot on activity 2 for the Health events study was represented by the correlation -0.35, but only two participants selected this correlation first as representing the scatter plot. However, on the corresponding activity for the Infant engagement study, seven of the participants selected the negative correlation first as the one that represented the target scatter plot.

In terms of students' understanding of correlations, participants' recorded responses to activity 3 are most revealing. On this activity for the Health events study, ten of the participants correctly arranged the correlations from the weakest to strongest relationship and four of the participants incorrectly positioned the positive correlation 0.60 as the strongest relationship. However, on activity 3 for the Infant engagement study, fourteen of the participants correctly arranged the set of correlations, but not one of the participants positioned the positive correlation 0.32 as the strongest relationship.

The McNemar tests for change described above indicated that students' misconceptions were not necessarily affected by their use of a relevant learner activity. However, qualitative analysis, which included an examination of participants' answers on the pre and post tests and their completion of the learner activities, revealed that in some cases students' misconceptions were affected by using the learner activities.

The case studies indicated that participants could find activity 1, which was designed to address a causalistic conception, confusing. The statistical issue of correlation and causality is subtle: students should not *just* infer that variable A is causally related to variable B, but they must also consider this possibility along with the three additional interpretations of a correlation. For some students activity 1 could be useful in this respect, but for other students they might think that they have to select one particular option or interpretation of a correlation.

Data from the case studies suggested that activity 2 is useful for students' learning because it links correlation coefficients with scatter plots. The findings from the case studies also suggested that more striking examples of correlations, such as correlations that are near -1 should be displayed by scatter plots if students are to fully appreciate the different kinds of relationships that can be obtained. Yet, relationships near -1 or 1 are not usually found in psychological research.

The findings of the summative study suggest that activity 3 could be used to address students' confusions about the strength of correlations. Here, students had the opportunity to arrange a set of correlations themselves and receive relevant feedback to their arrangement.

Drawing on theoretical perspectives outlined in chapter 4, *Link* uses real studies and genuine data from psychological research. It was thought that the context of a research study would facilitate the acquisition of statistical concepts. The case studies were partly set up to investigate the learning process while students worked with *Link*. It was hoped that qualitative data relating to what students thought about the studies in the program would be obtained. It was found, however, that the case study participants did not comment in particular on the studies provided in the program. It was therefore not possible to determine if the use of real studies and data contributed to students' learning about correlation.

9.7 Summary

This chapter described the final version of *Link* that was developed and a summative evaluation of this program. The evaluation study used the methodology detailed in chapter 6 and involved students from the University of Luton who were studying psychology. The study was designed to investigate whether *Link* contributed to students' general understanding of correlation and whether the learner activities in *Link* affected students' misconceptions. Quantitative findings relating to learning outcomes suggested that *Link*

successfully contributed to students' general understanding of correlation. Findings also indicated that the use of paper-based instructional materials, which were devised for the study, contributed to students' understanding of correlation. It was suggested that because the participants who took part in the study did not necessarily hold misconceptions about correlation, but lacked the necessary prior knowledge to answer questions or held additional confusions that did not relate to correlation *per se*, a form of instruction covering the topic of correlation, whether paper-based or computer-based, would have contributed to participants' general understanding of correlation. *Link* was specifically designed to address particular misconceptions about correlation that participants in the study did not necessarily hold.

Qualitative data analysis, which included case studies of participants using *Link*, revealed that in some cases the learner activities in *Link* addressed students' misconceptions. This meant that having used *Link*, participants' answers on the post test indicated that they had an appropriate understanding of a concept, rather than a misconception as indicated by their answers on the pre test.

Chapter 10

Conclusions

10.1 Introduction

This final chapter summarises the primary achievements of the research described in this thesis and the implications that this work has for research, education and for computer-assisted learning for statistics. Limitations of the thesis research are also outlined. Further improvements to the design of *Link* are considered and possible future research efforts are proposed.

10.2 Achievements

The research described in this thesis is concerned with psychology students' understanding of correlation and the design of a computer-assisted learning program. There have been a number of achievements in this research:

- The identification and comprehensive documentation of psychology students' misconceptions relating to correlation.
- The development of tests in correlation that provide an assessment of a student's understanding of the topic.
- A review of existing computer-assisted learning programs for correlation.
- The development of *Link*, a stand alone computer-assisted learning program.
- The use of psychological research studies and real data sets in *Link*.
- The formative, expert and summative evaluations of *Link*.

The research described in the earlier parts of the thesis indicated that psychology students are likely to find statistics difficult. The focus of this thesis was the important topic of correlation, specifically the kinds of difficulties and confusions that students

encounter in this area. With the increasing use of computer technology on statistics courses in higher education, it is likely that computer-assisted learning programs will be increasingly used as part of the statistics curriculum. A computer-assisted learning program could provide an additional form of instruction as part of this curriculum to help students acquire statistical concepts. However, it was argued in the thesis that the design of computer-assisted learning programs for statistics should be informed by research on students' learning, and research and developments in the field of computer-assisted learning. In addition, it is imperative that empirical work that informs the design of computer-assisted learning programs involves the target audience of students for a particular program, and that the evaluations of a program should involve this target audience as participants in the evaluation process.

The design of *Link* was informed by research-based principles of learning, an investigation that looked at students' difficulties relating to correlation and a review of computer-assisted learning programs. Drawing on theoretical perspectives (e.g., Bransford et al, 1990), it was proposed that the acquisition of statistical concepts is facilitated if they are presented to students in the context of a psychology study. In contrast to existing computer-assisted learning programs that cover correlation, *Link* provides two authentic psychological studies from the research literature and uses the real data from these studies. In addition, *Link* provides different kinds of learner activities that were designed to address students' misconceptions which have been identified through research (Morris, 1997).

It should be emphasised that psychology students must not only come to understand the concepts relating to correlation, but must also learn to interpret correlations in the context of psychological research. For example, students should be able to make sense of the findings from a study that employs a correlational design and be able to contemplate possible explanations of reported relationships between variables. Statistics textbooks aimed at psychology students typically present neat and tidy relationships,

which have coefficients of zero, -1, or 1, (or near such values) and are illustrated by scatter plots (e.g., Gravetter & Wallnau, 1995b, p. 343; Pagano, 1990, p. 119). Authentic data sets from research do not tend to be used in this context. In the case of computer-assisted learning programs that cover correlation, real data sets from the psychological research literature have not typically been used. (See chapter 4). However, as highlighted in chapter 4, constructivist approaches to learning have emphasised that concepts to be acquired by a learner should be presented in a realistic and meaningful context. Real data from psychological research can produce correlations that are near to 0.4 or 0.3, which if plotted give rather untidy scatter plots. In contrast to the majority of existing computer-assisted learning programs, *Link* makes use of data sets from authentic research studies and therefore provides learners with a variety of different kinds of real relationships.

The development of *Link* was informed by a formative evaluation study, which consisted of two phases, and an expert evaluation. The empirical work described in this thesis involved psychology students studying at a variety of institutions: full time students at the universities of Buckingham and Luton and part time students who were studying with the Open University.

Research that was discussed in chapter 2 of the thesis indicated that students are likely to find statistical concepts difficult to acquire, but there has been little research that has focused on psychology students' understanding of correlation. A study, which was described in chapter 3, was conducted to investigate students' difficulties and confusions concerning correlation. The findings of this investigation indicated that students held confusions relating to negative correlations, the strength of correlations and the statistical issue of causation and correlation. These findings were discussed with regard to other research (Batanero et al, 1997) and three misconceptions were outlined: causalistic conception, unidirectional conception and the conception that a positive correlation is stronger than a negative correlation. The formative and summative evaluation studies of

Link provided data confirming that students held these misconceptions. Findings from these studies also showed that students held confusions relating to levels of statistical significance where a student would, for example, think that a correlation coefficient indicated a significance level.

An outcome of the thesis research was the development of tests that were designed to assess students' general understanding of correlation and to identify misconceptions. These tests used questions that were piloted in the investigation that was described in chapter 3, and were modified for the formative and summative evaluation studies. In these studies, the tests were used as pre and post tests so that learning outcomes could be assessed in the evaluation of *Link*.

The design of *Link* was informed by a review of computer-assisted learning programs that cover correlation. Related work has involved the evaluation of computer-based learning materials for statistics for a masters course on Advanced Experimental Design and Analysis in psychology (Morris & Le Voi, 1997, 1998). The design of *Link* was also informed by a formative study that was designed to assess the usability and possible instructional effectiveness of earlier prototypes. This study employed a methodology for the evaluation of computer-assisted learning that was based on an existing framework (Jones et al, 1996). The formative evaluation provided data relating to learning outcomes, the process of learning while students used *Link* and also data concerning students' opinions of the program. In terms of learning outcomes, it was found that there was a significant increase in students' scores from the pre test to the post test, which suggested that *Link* contributed to students' general understanding of correlation (Morris, 1998b). The findings of the formative study were used to improve the design of the learner activities and related material provided by *Link*. The program was also modified with regard to an expert evaluation, which suggested that feedback to a learner activity would need to be changed so that it is contingent on a user's response to an activity.

The final version of *Link* was developed to provide two studies from psychological research. The genuine data sets from these studies were used to provide correlations and scatter plots that are presented in the program. This meant that the program provided two studies and six learner activities that were designed to address students' misconceptions relating to correlation. The evaluative framework described in chapter 6 was employed in a summative evaluation study that used a pre-test-post-test control group design. This study was designed to investigate whether *Link* contributed to students' general understanding of correlation and whether particular learner activities in *Link* affected students' misconceptions. Findings of the summative study showed that having used *Link*, students' scores significantly increased from the pre to the post test, indicating that the program contributed to students' general understanding of correlation. However, this increase was also found in the instructional control group, where students completed paper-based instructional materials covering the topic of correlation, but not in the basic control group.

The above finding of the summative evaluation was interpreted as follows. It was found that the participants who took part in the study held particular misconceptions, which *Link* was designed to address, but also lacked the necessary knowledge to answer questions on the pre test, or held additional confusions that were not specifically designed to be addressed by the program. It was suggested that learning materials that provide an introduction to correlation, whether these be paper-based or computer-based, would have contributed to the participants' understanding of correlation. It was proposed that a pre test could be used to identify students' misconceptions relating to correlation and *Link* could therefore be used appropriately if students held the misconceptions that the learner activities in *Link* were specifically designed to address.

The above finding of the summative study was also discussed in the light of a similar research outcome (Shute & Gawlick-Grendell, 1994), and it was suggested that the findings are encouraging given that the students who took part in the study were not

familiar with using computer-assisted learning programs for statistics.

The findings of the summative evaluation study indicated that the learner activities in *Link* did not necessarily affect students' misconceptions in correlation. However, this study provided further qualitative data relating to students' misconceptions. It was found, for example, that students held confusions relating to causality, negative correlations and the strength of correlations, but that they also held confusions pertaining to levels of statistical significance that are used in psychological research.

One of the major outcomes of the thesis research was the design and development of a computer-assisted learning program. This meant that a substantial amount of research work involved program design, implementation and testing that was carried out by the thesis author. The result of this work was a Macromedia Director application called *Link* that was designed to be used by psychology students in higher education.

10.3 Implications for research

The empirical work that was described in this thesis suggests that students studying psychology have a variety of misconceptions relating to correlation. This empirical work identified the following misconceptions: unidirectional conception, causalistic conception and the conception that a positive correlation is stronger than a negative correlation. This finding is not entirely consistent with research that has been carried out in Spain, which has looked at students' understanding of association in general (Batanero et al, 1997). Batanero and her colleagues (1997) have reported that students hold a unidirectional conception of correlation and a causalistic conception of correlation, but they have not found that students have confusions about the strength of correlations. Students who participated in the studies, which were described in this thesis, held confusions relating to the strength of correlations and would think that a positive correlation is stronger than a negative correlation. In addition, this work found that students' held additional confusions not directly relating to correlation, where a student would, for example,

confuse correlation coefficients and levels of statistical significance.

It is therefore suggested that further work is required to investigate students' understanding of statistical significance and how this understanding might impact on students' understanding of correlation. For example, a student might interpret a correlation that is reported to be significant differently from one that is not, and this may have implications for students' understanding of causality or the strength of correlations.

Much research on students' learning has focused on students' misconceptions in a variety of subject matter areas. The findings of this thesis concerning students' misconceptions indicates that students not only hold misconceptions that directly relate to the subject area under investigation, but also hold additional confusions. It was found that the psychology students who took part in the summative study were confused by correlation coefficients and levels of statistical significance. This was in spite of the fact that the pre and post tests used in the study did not include the words significance or probability.

This finding indicates that even though research tends to focus on particular subject matter areas, students learn topics in the context of a wider subject area. It is likely that researchers have attempted to identify students' misconceptions relating to specific topic areas because this approach limits the field of enquiry. However, it is suggested that topic areas are understood by students as part of an integrated subject matter area. For example, in the case of the topic of correlation, students should learn that a correlation is a measure of association (among many) and that it can be used as a descriptive or as an inferential statistic. Inevitably, this means that students' understanding of correlation should be part of their wider understanding of the use of statistics in psychology. This implies that research needs to focus on students' understanding of particular topic areas, but must also look at the students' understanding of that topic in relation to the broader subject area under consideration.

The findings of the summative evaluation indicated that *Link* did not necessarily affect students' misconceptions in correlation, which the program was designed to address. It is noteworthy that the paper-based instructional materials that were devised for the study also did not necessarily affect students' misconceptions. However, this does not mean that *Link* should not be used to shift students' misconceptions. It was clear from the findings of the summative study that participants held particular misconceptions (e.g., causalistic), but also lacked the necessary prior knowledge or held additional confusions. In the light of these findings, the research described in this thesis does not necessarily indicate that students' statistical misconceptions are persistent in spite of instructional efforts.

10.4 Implications for computer-assisted learning for statistics

The findings of this thesis indicate that computer-assisted learning programs can be designed with students' confusions in mind. A program for correlation should cover the concepts of positive, negative and zero correlations, the strength of correlations and correlation and causation. In addition, a correlation is a descriptive measure and an inferential statistic and this must be made explicit to a learner to avoid the confusion that students have about correlation coefficients and levels of significance.

The summative evaluation study in this thesis found that students' general understanding of correlation was improved whether they used *Link* or paper-based instructional materials. This outcome was interpreted above in the light of the findings that showed that students held misconceptions about correlation, but also lacked prior knowledge or held additional misconceptions. However, the outcome that students' general understanding was improved irrespective of the instructional medium has two primary implications for computer-assisted learning for statistics. Firstly, the use of computer-assisted learning programs instead of paper-based instructional materials should be carefully considered. In the case of *Link*, it could be used for students who hold particular misconceptions about correlation, which could be identified though the use of a

diagnostic test. Secondly, if computer-assisted learning programs are used in higher education, then the instructional capabilities of the computer should be harnessed. This would mean that a program would provide an additional and *alternative* form of instruction that might help certain students to acquire statistical concepts.

10.5 Implications for education

Teachers of statistics for psychology students in higher education should be aware that students can have difficulties and confusions relating to correlation: students can be confused by negative correlations, the strength of correlations and causality, and they might also interpret a coefficient as a level of significance. In addition, the first study that was described in this thesis showed that students have difficulties in deriving a procedure from a formula to obtain a statistic and in interpreting correlations in the context of a research study. Teachers of statistics must emphasise the difference between positive, negative and zero correlation and must demonstrate that a negative correlation does indicate a relationship that is inverse. Students should be taught that data in a scatter plot is represented by a single value of a correlation coefficient and that different correlation coefficients are computed depending on the level of measurement used for data. A teacher should emphasise that a correlation describes a relationship between two sets of variables, but is also a statistic that can be tested to see if it is significant.

Teachers of statistics must think very carefully about whether it is necessary for students to learn to calculate statistics by following predefined procedures, rather than ensuring that students learn how to derive a computational procedure from a formula so that they can obtain a statistic. The statistics curriculum is changing with the increasing availability of data-analysis software, and teachers of statistics should make sure that students learn how to use these packages appropriately and be able to interpret the output that is generated by such applications.

10.6 Limitations of the research

The investigation described in chapter 3, which looked at psychology students' difficulties with correlation, focused on positive, negative and zero correlations, the strength of correlations, the calculation of a correlation, the significance test for a correlation, causality and the interpretation of correlations. This study and the subsequent thesis work did not look at how students might understand the effect that outliers have on the value of a correlation coefficient. In addition, there are patterns that show up on scatter plots, such as curvilinear relationships which give a correlation coefficient of zero. Students' understanding of these two issues was not addressed in the research because it was beyond the scope of the thesis, but could be considered in further research as outlined below. (See 10.8 Further research).

The design of *Link* was based on research-based principles of learning and on empirical work that identified students' misconceptions about correlation. Accordingly, computer-based learner activities were designed to address particular misconceptions. This can be described as a principled approach to design because students' prior knowledge was seriously considered. However, there is a primary problem with this approach: it is very difficult to consider and therefore anticipate all conceivable errors or confusions that students might hold about a particular topic area. The findings of the summative evaluation highlighted this problem. It was found that although learner activities in *Link* had been carefully designed to address the confusions that students hold about negative correlations, the strength of correlations and causality, students who took part in the summative study held additional confusions relating to, for example, statistical significance.

In the formative study, all of the students were observed while they worked though *Link* and thought aloud. This set up could have influenced the learning process and may not represent the usual conditions where students learn about correlation. Similarly, the

setting for the summative study was not naturalistic because participants were paid for their time and worked though *Link* in a cubicle in the psychology laboratory. Students who take part in a study under these conditions might not have the same motivation to learn about correlation as, say, a group of students who attend a computer laboratory class set up for their statistics course (c.f. Draper et al, 1996).

Link uses authentic studies from the research literature and successfully contributed to students' general understanding of correlation. However, it was not possible to determine from the findings of the summative evaluation study whether the use of studies specifically contributed to students' learning about correlation. This was because the participants whose interactions were used to illustrate the use of the program did not, for example, comment on the use of the studies in the program.

10.7 Further improvements to *Link*

The use of *Link* successfully contributed to students' general understanding of correlation. However, the findings of the summative study suggested that activity 1 did not always influence students' causalistic conceptions. The development of *Link* was described at a Computers in Psychology Conference (Morris, 1998a), and feedback from delegates referred to how activity 1 might confuse a learner. Activity 1 was designed so that a learner would be aware that there are a variety of interpretations to a correlation coefficient. The activity does not sufficiently emphasise that when interpreting the meaning of a correlation, a causal explanation can be ruled out because there are other explanations of the relationship where, for example, an additional variable C is responsible for the obtained correlation between A and B. To improve *Link*, activity 1 should be re-designed, but further work is needed to see how the statistical issue of causation and correlation can be addressed through instruction. It was thought that examples used in activity 1, in which a correlation is obtained between variables, such as brain size and IQ, would suggest to a learner that a causal relationship could not exist between variables and that they would therefore have to entertain other explanations.

Further work is required on the role of examples and how their use might help students appreciate the issue of causality in the interpretation of a correlation.

The findings of the summative evaluation suggest that activity 2 should be modified to provide scatter plots that represent correlations that are near to -1 or 1 and which would therefore provide clearly defined relationships to a learner. Indeed, statistical texts for the behavioural sciences often present a set of scatter plots that show tidy patterns and perfect relationships (e.g., Coolican, 1990; Pagano, 1990). However, as discussed above, data sets from research do not necessarily provide scatter plots with tidy clusters of plots or correlations that are near to perfect relationships and psychology students need to learn to interpret data and associated statistics that are presented in the context of research. To understand correlations, it is likely that students need to view a variety of different kinds of relationships displayed on scatter plots. Activity 2 could be re-designed to include scatter plots that show relationships that are near to -1 and 1, but this would mean that the genuine data sets used in *Link* could not be used for this purpose.

10.8 Further research

Link can be developed further as described above and this would mean that additional empirical work could involve the evaluation of a revised program. Additional research could focus on those students who hold the particular misconceptions that *Link* was designed to address. This research could therefore use diagnostic tests to identify students' misconceptions about correlation, such as a unidirectional conception. Further research could also be conducted to see how a unidirectional conception of correlation is related to the conception that a positive correlation is stronger than a negative correlation. A study could be run to focus on those students who have confusions relating to negative correlations and the strength of correlations. These students could be interviewed prior to using *Link* so that a detailed record of their understanding of correlations could be collected, and observed and asked to think aloud while they use the program so that data concerning the learning process could be recorded. A picture of how students come to

understand correlations by using computer-based learner activities could then be detailed.

Further research should be conducted to investigate students' understanding of levels of significance and how this understanding impacts on students' interpretation of correlation coefficients. The tests in correlation that were devised for the thesis research could be developed to include questions that relate to correlations that are reported as statistically significant. These questions could be used to identify students' misconceptions relating to significance.

Further research is required to investigate the use of examples in computer-based learning activities that are designed to address the statistical issue of correlation and causality. It is likely that particular examples, such as the relationship between TV violence and aggression prompt students to think about why the relationship is found to exist and whether other variables are responsible for the obtained correlation. There might be other examples of relationships between variables that do not prompt students to think in this way.

It was noted above that students' understanding of how outliers in a data set can influence the value of a correlation was not investigated in this thesis. It is noteworthy that the program *Understanding Statistics*, which was reviewed in chapter 4, provides scatter plots that have outliers. When plotted, bivariate data can show a curvilinear relationship, but the Pearson correlation coefficient, for example, will give a value of zero for the data, indicating no relationship. Students' understanding of the effect that outliers and non-linear relationships have on a correlation could be investigated if additional data sets were used in *Link*. This will of course mean that questions relating to these issues would have to be included in the tests in correlation that were developed for the thesis.

10.9 Summary

This final chapter has outlined the main achievements of the thesis research. These included: the identification and documentation of students' misconceptions about correlation, the development of tests in correlation, a review of computer-assisted learning programs for statistics, and the design and evaluation of *Link*. Implications that this thesis work has for research, education and for computer-assisted learning programs were described. Improvements to the design of *Link* were outlined and further empirical work was suggested. Further research could include an investigation of students' understanding of levels of statistical significance and how this relates to their interpretation of correlations. Empirical work could also investigate students' understanding of how outliers can influence the value of a correlation.

References

Anscombe, F. J. (1973) Graphs in Statistical Analysis. *The American Statistician*, 27, 1, 17 - 21.

Bangert-Drowns, R. L. & Kozma, R. B. (1989) Assessing the Design of Instructional Software. *Journal of Research on Computing in Education*, 21, 3, 241 - 262.

Barnett, V. & Holmes, P. (1996) *Understanding Statistics. User Manual.* Royal Statistical Society.

Batanero, C., Estepa, A., Godino, J. D. & Green, D. R. (1996) Intuitive Strategies and Preconceptions about Association in Contingency Tables. *Journal for Research in Mathematics Education,* 27, 2, 151 - 169.

Batanero, C., Estepa, A. & Godino, J. D. (1997) Evolution of students' understanding of statistical association in a computer based teaching environment. In J. B. Garfield and G. Burrill (Eds.), *Research on the Role of Technology in Teaching and Learning Statistics. Proceedings of the 1996 IASE Round Table Conference.* (pp. 191 - 205). Voorburg, The Netherlands: International Statistical Institute.

Benson, J. B. (1993) Season of Birth and Onset of Locomotion: Theoretical and Methodological Implications. *Infant Behavior and Development*, 16, 69 - 81.

Biehler, R. (1994) Cognitive technologies for statistics education: Relating the perspective of tools for learning and tools for doing statistics. In L. Brunelli & G. Cicchitelli (Eds.), *Proceedings of the First Scientific Meeting of the International Association for Statistical Education.* (pp. 173 - 190). Perugia: Universita di Perugia.

Bransford, J. D., Sherwood, R. D., Hasselbring, T. S., Kinzer, C. K. & Williams, S. M. (1990) Anchored Instruction: Why We Need It and How Technology

Can Help. In P. Nix and R. Spiro (Eds.), *Cognition, Education and Multimedia.* (pp. 115 - 141). Hillsdale, NJ: Erlbaum.

Brewer, J. K. (1985) Behavioral statistics textbooks: Source of myths and misconceptions? *Journal of Educational Statistics,* 10, 3, 252 - 268.

Brown, J. S., Collins, A. & Duguid, P. (1989) Situated cognition and the culture of learning. *Educational Researcher*, 18, 1, 32 - 42.

Chapman, L. J. (1967) Illusory correlation in observational report. *Journal of Verbal Learning and Behavior,* 6, 151 -155.

Chapman, L. J. & Chapman, J. (1982) Test results are what you think they are. In D. Kahnemann, P. Slovic and A. Tversky (Eds.), *Judgment Under Uncertainty: Heuristics and Biases.* (pp. 239 - 248). Cambridge: Cambridge University Press.

Choi, J. & Hannafin, M. (1995) Situated Cognition and Learning Environments: Roles, Structures, and Implications for Design. *Educational Technology Research and Development,* 43, 2, 53 - 69.

Clark-Carter, D. (1997) *Doing Quantitative Psychological Research: From Design to Report.* Psychology Press: Hove, UK.

Cognition and Technology Group (1992) An Anchored Instruction Approach to Cognitive Skills Acquisition and Intelligent Tutoring. In J. W. Regian and V. J. Shute (Eds.), *Cognitive Approaches to Automated Instruction* (pp. 135 - 170). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Cohen, G. & Java, R. (1995) Memory for Medical History: Accuracy of Recall. *Applied Cognitive Psychology*, 9, 273 - 288.

Coolican, H. (1990) *Research Methods and Statistics in Psychology.* London: Hodder & Stoughton.

Cumming, G. & Thomason, N. (1995) Designing software for cognitive change: StatPlay and understanding statistics. In J. D. Tinsley and T. J. van Weerts (Eds.), *World Conference on Computers in Education VI. WCCE 95, Liberating the Learner.* (pp. 753 - 765). London: Chapman and Hall.

Cumming, G. & Thomason, N. (1998) StatPlay: Multimedia for statistical understanding. In L. Pereira-Mendoza, L. Seu Kea, T. Wee Kee & W. Wong (Eds.), *Statistical Education - Expanding the Network. Proceedings of the Fifth International Conference on Teaching of Statistics. Volume 2.* (pp. 947 - 952). 21 - 26 June 1998 Singapore. International Statistical Institute.

Daly, F., Hand, D. J., Jones, M. C., Lunn, A. D. & McConway, K. J. (1995) *Elements of Statistics.* Wokingham, England: Addison-Wesley Publishing Company.

De Corte, E. (1995) Learning Theory and Instructional Science. In P. Reimann and H. Spada (Eds.), *Learning in Humans and Machines: Towards an Interdisciplinary Learning Science.* (pp. 97 - 108). Oxford: Pergamon.

Dix, A., Finlay, J., Abowd, G, & Beale, R. (1993) *Human Computer Interaction.* New York: Prentice Hall.

Draper, S. W., Brown, M. I., Edgerton, E., Henderson, F. P., McAteer, E., Smith, E. D. & Watt, H. D. (1994) *Observing and Measuring The Performance of Educational Technology.* A report by the University of Glasgow's institutional project in the Teaching and Learning Technology Programme *Teaching with Independent Learning Technologies.* The University of Glasgow.

Draper, S. W., Brown, M. I., Henderson, F. P. & McAteer, E. (1996) Integrative Evaluation: An Emerging Role for Classroom Studies of CAL. *Computers and Education*, 26, 1 - 3, 17 - 32.

Driver, R. (1988) Theory into Practice II: A Constructivist Approach to Curriculum Development. In P. Fensham (Ed.), *Development and Dilemmas in Science Education*. (pp. 133 - 149). London: The Falmer Press.

du Boulay, B. (1998) What does the "AI" in AIED buy? *Proceedings of the Colloquim on Artificial Intelligence in Educational Software, IEE Digest no. 98/313.*

Dugard, P. & Todman, J. (1995) Analysis of Pre-test-Post-test Control Group Designs in Educational Research. *Educational Psychology*, 15, 2, 181 - 198.

Ericsson, K. A. & Simon, H. A (1980) Verbal Reports as Data. *Psychological Review*, 87, 3, 215 - 251.

Ericsson, K. A. & Simon, H. A (1993) *Protocol Analysis. Verbal Reports as Data*. Revised Edition. Cambridge, Massachusetts: The MIT Press.

Eron, L. D., Huesmann, L. R., Lefkowitz, M. M & Walder, L. O. (1972) Does television violence cause aggression? *American Psychologist,* 27, 253 - 263.

Estepa, A. & Batanero, C. (1996) Judgments of correlation in scatter plots: Students' intuitive strategies and preconceptions. *Hiroshima Journal of Mathematics Education,* 4, 25 - 41.

Finch, S. & Cumming, G. (1998) Assessing conceptual change in learning statistics. In L. Pereira-Mendoza, L. Seu Kea, T. Wee Kee & W. Wong (Eds.), *Statistical Education - Expanding the Network. Proceedings of the Fifth International Conference on Teaching of Statistics. Volume 2.* (pp. 897 - 903). 21 - 26 June 1998 Singapore. International Statistical Institute.

Fox, W. P. & Fowler, C. W. (1996) Aiding undergraduate understanding of covariance and correlation. *PRIMUS*, 6, 3, 235 - 243.

Gagne, R. M., Briggs, L. J. & Wager, W. W. (1988) *Principles of Instructional Design.* New York: Holt, Rinehart and Winston.

Garfield, J. & 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.

Goldstein, M. D. & Strube, M. J. (1995) Understanding Correlations: Two Computer Exercises. *Teaching of Psychology*, 22, 3, 205 - 206.

Gravetter, F. J. & Wallnau, L. B. (1995a) *Instructor's Manual to Accompany Essentials of Statistics for the Behavioral Sciences.* Second Edition. New York: West Publishing Company.

Gravetter, F. J. & Wallnau, L. B. (1995b) *Essentials of Statistics for the Behavioral Sciences.* Second Edition. New York: West Publishing Company.

Greene, J. & d'Oliveira, M. (1982) *Learning to Use Statistical Tests in Psychology: A Student's Guide.* Milton Keynes: The Open University Press.

Hand, D. J., Daly, F., Lunn, A. D., McConway, K. J. & Ostrowski, E. (Eds.) (1994) *A Handbook of Small Data Sets.* London: Chapman & Hall.

Hardiman, P. T., Well, A. D. & Pollatsek, A. (1984) Usefulness of a Balance Model in Understanding the Mean. *Journal of Educational Psychology,* 76, 5, 792 - 801.

Hawkins, A., Jolliffe, F. & Glickman, L. (1992) *Teaching Statistical Concepts.* London: Longman.

Hennessy, S., Twigger, D., Driver, R., O'Shea, T., O'Malley, C. E., Byard, M., Draper, S., Hartley, R., Mohamed, R. & Scanlon, E. (1995) Design of a computer-augmented curriculum for mechanics. *International Journal of Science Education*, 17, 1, 75 - 92.

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

Hix, D. & Hartson, H. R. (1993) *Developing User Interfaces. Ensuring Usability Through Product & Process.* New York: John Wiley & Sons, Inc.

Howell, D. C. (1992) *Statistical Methods for Psychology.* Third Edition. Belmont, California: Duxbury Press.

Jennings, D. L., Amabile, T. M. & Ross, L. (1982) Informal covariation assessment: Data-based versus theory-based judgments. In D. Kahneman, P. Slovic, and A. Tversky (Eds.), *Judgment Under Uncertainty: Heuristics and Biases.* (pp. 211 - 230). Cambridge: Cambridge University Press.

Jones, A., Scanlon, E., Tosunoglu, C., Ross, S., Butcher, P., Murphy, P. & Greenberg, J. (1996) Evaluating CAL at the Open University: 15 Years On. *Computers and Education*, 26, 1 - 3, 5 - 15.

Kahneman, D. & Tversky, A. (1972) Subjective Probability: A Judgment of Representativeness. *Cognitive Psychology,* 3, 430 - 454.

Konold, C. (1989) Informal conceptions of probability. *Cognition and Instruction,* 6, 1, 59 - 98.

Lane, D. M., Anderson, C. A. & Kellam, K. L. (1985) Judging the relatedness of variables: The psychophysics of covariation detection. *Journal of Experimental Psychology: Human Perception and Performance,* 11, 5, 640 - 649.

Laurillard, D. (1993) *Rethinking University Teaching. A Framework for the Effective Use of Educational Technology.* London: Routledge.

Lefkowitz, M. M., Eron, L. D., Walder, L. O. & Huesmann, L. R. (1977) *Growing Up to be Violent.* New York: Pergamon.

Marriott, F. H. C. (1990) *A Dictionary of Statistical Terms.* Fifth Edition. Harlow, England: Published for the International Statistical Institute by Longman Scientific and Technical.

Mevarech, Z. R. (1983) A deep structure model of students' statistical misconceptions. *Educational Studies in Mathematics*, 14, 415 - 429.

Minium, E. W. (1978) *Statistical Reasoning in Psychology and Education.* Second Edition. New York: John Wiley & Sons.

Milheim, W. D. (1995 - 1996) Interactivity and Computer-Based Instruction. *Journal of Educational Technology Systems*, 24, 3, 225 - 233.

Monk, A., Wright, P., Haber, J. & Davenport, L. (1993) *Improving Your Human-Computer Interface: A Practical Technique.* New York: Prentice-Hall.

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. The Open University.

Morris, E. J. (1998a) The development of a computer-assisted learning program designed to address students' misconceptions in correlation. *CiP98 Conference Proceedings. Computers in Psychology Conference.* 15 - 17 April 1998. (pp. 55 - 56). University of York.

Morris, E. J. (1998b) *Link*: The principled design of a computer-assisted learning program for correlation. In L. Pereira-Mendoza, L. Seu Kea, T. Wee Kee and W. Wong (Eds.), *Statistical Education - Expanding the Network. Proceedings of the Fifth International Conference on Teaching Statistics. Volume 2.* Singapore, 21 - 26 June, 1998. (pp. 1033 - 1039). International Statistical Institute.

Morris, E. J. & Le Voi, M. (1997) Using computers and web-based statistical resources. *Department of Psychology Report Series*. The Open University.

Morris, E. J. & Le Voi, M. (1998) Sorting the wheat from the chaff: evaluating a large corpus of computer-based statistics teaching resources. In L. Pereira-Mendoza, L. Seu Kea, T. Wee Kee and W. Wong (Eds.), *Statistical Education - Expanding the Network. Proceedings of the Fifth International Conference on Teaching Statistics. Volume 2*. Singapore, 21 - 26 June, 1998. (pp. 977 - 983). International Statistical Institute.

Oates, J. (1998) Risk Factors in Infant Attrition and Low Engagement in Experiments and Free-Play. *Infant Behavior and Development*, 21, 4, 555 - 569.

Ohlsson, S. (1991) Knowledge Requirements for Teaching: The Case of Fractions. In P. Goodyear (Ed.), *Teaching Knowledge and Intelligent Tutoring*. (pp. 25 - 59). New Jersey: Albex Publishing Corporation.

Ohlsson, S. (1993) Impact of Cognitive Theory on the Practice of Courseware Authoring. *Journal of Computer Assisted Learning*, 9, 194 - 221.

O'Shea, T. (1992) Educational Technology Works. British Association for Advancement of Science 1992 Presidential Lecture - Psychology. *CITE Report No. 166*. Centre for Information Technology in Education. Institute of Educational Technology. The Open University.

Pagano, R. R. (1990) *Understanding Statistics in the Behavioral Sciences*. Third Edition. New York: West Publishing Company.

Perkins, D. N. & Martin, F. (1986) Fragile Knowledge and Neglected Strategies in Novice Programmers. In E. Soloway & S. Iyengar (Eds.), *Empirical Studies of Programmers. Papers presented at the First Workshop on Empirical Studies of*

Programmers, June 5 - 6, 1986, Washington, DC. (pp. 213 - 229). Norwood, New Jersey: Ablex Publishing Corporation.

Persidsky, A. (1996) *Director 5 for Macintosh.* Berkeley, US: Peachpit Press.

Pervin, L. A. (1989) *Personality. Theory and Research.* Fifth Edition. New York: John Wiley & Sons, Inc.

Pierce, R. & Roberts, L. (1998) Introductory statistics: Critical evaluation and clear communication. In L. Pereira-Mendoza, L. Seu Kea, T. Wee Kee & W. Wong (Eds.), *Statistical Education - Expanding the Network. Proceedings of the Fifth International Conference on Teaching of Statistics. Volume 3.* (pp. 1201 -1206). 21 - 26 June 1998 Singapore. International Statistical Institute.

Pollatsek, A., Lima, S. & Well, A. D. (1981) Concept or Computation: Students' Understanding of the Mean. *Educational Studies in Mathematics,* 12, 191 - 204.

Preece, J. (Ed.) (1993) *A guide to usability. Human factors in computing.* Wokingham, England: Addison-Wesley.

Reiser, R. A. & Kegelmann, H. W. (1994) Evaluating Instructional Software: A Review and Critique of Current Methods. *Educational Technology Research and Development*, 42, 3, 63 - 69.

Savery, J. R. & Duffy, T. M. (1995) Problem Based Learning: An Instructional Model and Its Constructivist Framework. *Educational Technology*, 35, 5, 31 - 38.

Shaughnessy, J. M. (1992) Research in Probability and Statistics: Reflections and Directions. In D. A. Grouws (Ed.), *Handbook of Research on Mathematics Teaching and Learning. A project of the National Council of Teachers of Mathematics.* (pp. 465 - 494). New York: MacMillan Publishing Company.

Shavelson, J. (1981) *Statistical Reasoning for the Behavioral Sciences.* London: Allyn and Bacon, Inc.

Shuell, T. J. (1992) Designing Instructional Computing Systems for Meaningful Learning. In M. Jones and P. H. Winne (Eds.), *Adaptive Learning Environments.* NATO ASI Series, Vol. F 85. (pp. 19 - 53). Berlin Heidelberg: Springer-Verlag.

Shute, V. J. & Gawlick-Grendell, L. A. (1993) An Experiential Approach to Teaching and Learning Probability: Stat Lady. In P. Brna, S. Ohlsson & H. Pain (Eds.), *Artificial Intelligence in Education. Proceedings of AIED 93, World Conference on Artificial Intelligence in Education. Edinburgh, August.* (pp. 177 - 184). Charlottesville, VA: Association for the Advancement of Computing in Education.

Shute, V. J. & Gawlick-Grendell, L. A. (1994) What does the computer contribute to learning? *Computers and Education*, 23, 3, 177 - 186.

Shute, V. J., Gawlick-Grendell, L. A., Young, R. K. & Burnham, C. A. (1996) An experiential system for learning probability: Stat Lady description and evaluation. *Instructional Science*, 24, 25 - 46.

Siegel, S. & Castellan, N. J. (1988) *Non Parametric Statistics in the Behavioral Sciences.* Second Edition. New York: McGraw-Hill.

Smith, G. E., Cohen, S., Brown, L. C., Chechile, R. A., Garman, D., Cook, R. G., Ennis, J. G. & Lewis, S. (1997) *ConStatS: Software for Conceptualizing Statistics. A User's Manual.* New Jersey: Prentice Hall.

Smith, P. K. & Cowie, H. (1988) *Understanding Children's Development.* Oxford: Basil Blackwell.

Smith, J. P., diSessa, A. A. & Roschelle, J. (1993) Misconceptions Reconceived: A Constructivist Analysis of Knowledge in Transition. *The Journal of the Learning Sciences*, 3, 2, 115 - 163.

Songer, C. J. & Mintzes, J. J. (1994) Understanding Cellular Respiration: An Analysis of Conceptual Change in College Biology. *Journal of Research in Science Teaching*, 31, 6, 621 - 637.

Strahan, R. F. & Hansen, C. J. (1978) Underestimating correlation from scatterplots. *Applied Psychological Measurement,* 2, 4, 543 - 550.

Taylor, R. P. (1980) Introduction. In R. P. Taylor (Ed.), *The computer in the school: Tutor, tool, tutee.* (pp. 1 - 10). New York: Teachers College Press.

The Open University (1990) Social Sciences/Science/School of Education. An Interfaculty Second Level Course. Introduction to Psychology. DSE 202. Methodology Handbook. Part 1. *Correlational Research Designs.* Prepared on behalf of the Course Team by Judith Greene. Milton Keynes: The Open University.

Thomason, N., Cumming, G. & Zangari, M. (1994) Understanding central concepts of statistics and experimental design in the social sciences. In K. Beattie, C. McNaught, and Wills (Eds.), *Interactive Multimedia in University Education: Designing for Change in Teaching and Learning.* (pp. 59 - 81). Elsevier Science B. V. (North-Holland).

Tversky, A. & Kahneman, D. (1982) Judgment under uncertainty: Heuristics and biases. In D. Kahneman, P. Slovic, and A. Tversky (Eds.), *Judgment Under Uncertainty: Heuristics and Biases.* (pp. 3 - 20). Cambridge: Cambridge University Press.

Tversky, A. & Kahneman, D. (1993) Belief in the Law of Small Numbers. In G. Keren and C. Lewis (Eds.), *A Handbook for Data Analysis in the Behavioural Sciences: Methodological Issues.* (pp. 341 - 349). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Wagner, B. M., Compas, B. E. & Howell, D. C. (1988) Daily and Major Life Events: A Test of an Integrative Model of Psychosocial Stress. *American Journal of Community Psychology*, 16, 2, 189 - 205.

Weinberg, S. L. & Goldberg, K. P. (1990) *Statistics for the Behavioral Sciences.* Cambridge: Cambridge University Press.

Well, A. D., Boyce, S. J., Morris, R. K., Shinjo, M. & Chumbley, J. I. (1988) Prediction and judgment as indicators of sensitivity to covariation of continuous variables. *Memory & Cognition,* 16, 3, 271 - 280.

Wenger, E. (1987) *Artificial Intelligence and Tutoring Systems. Computational and Cognitive Approaches to the Communication of Knowledge.* Los Altos, CA: Morgan Kaufmann.

Willerman, L. Schultz, R., Rutledge, J. N. & Bigler, E. (1991) In Vivo Brain Size and Intelligence. *Intelligence,* 15, 223 - 228.

Zahner, J. E., Reiser, R. A., Dick, W. & Gill, B. (1992) Evaluating Instructional Software: A Simplified Model. *Educational Technology Research and Development*, 40, 3, 55 - 62.

Appendix A

Information sheet for study

Thank you for agreeing to take part in this study

The purpose of this study is to look at students' understanding and skills in the topic correlation. The session that follows is not a test of your statistical knowledge. You are ensured confidentiality in the responses you give, or the answers and explanations you provide to the questions that have been set. The study will involve the following:

• I will ask you to complete a questionnaire.

• I will provide you with a Student task booklet that contains questions in statistics that I would like you to answer. Answer and work through these questions at your own pace. For the questions there is a calculator available and there is plenty of space in the booklet to write your answers. Could you please also show your working in the space provided. Whilst you work through the questions I would like you to think aloud. This technique is very useful because it gives me a clearer idea of how, for example, you might have reached a particular answer. To think aloud whilst you work through the questions simply say what you are thinking out aloud. It might feel slightly unusual at first, but there will be a chance to practice before you start on the booklet.

• Whilst you are working through the Student task booklet I shall be observing and taking notes, but please remember I am not testing you. If you get very stuck on a particular question I will follow a series of steps to help you to continue.

• If you have any questions that come to mind whilst you work through the questions, please save them for the end of the session when I can talk to you about them.

• When you have completed all the questions in the Student task booklet I will ask you some follow-up questions.

• The whole of the session will be recorded on audio cassette and will take a maximum of 90 minutes.

If you have any questions, please ask them now.

Appendix B

Questions for task booklet (including model answers)

Correlational designs

Question 1

In psychology, when would you use a correlational design?

Answer

If you wanted to see if there was a relationship or association between two variables.

Correlational designs

Question 2

Give an example of a study that would make use of a correlational design.

Answer

An example of a study that looks at the relationship between two variables. (For example, spelling and reading ability; students' scores on two examinations, etc.)

Correlational designs

Question 3

Let's suppose that a large-scale research study has reported that a significant correlation had been found between clinical depression and cancer. What do the findings tell us about the statistical relationship between clinical depression and cancer?

Appendix B

Questions for task booklet (including model answers)

Correlational designs

Question 1

In psychology, when would you use a correlational design?

Answer

If you wanted to see if there was a relationship or association between two variables.

Correlational designs

Question 2

Give an example of a study that would make use of a correlational design.

Answer

An example of a study that looks at the relationship between two variables. (For example, spelling and reading ability; students' scores on two examinations, etc.)

Correlational designs

Question 3

Let's suppose that a large-scale research study has reported that a significant correlation had been found between clinical depression and cancer. What do the findings tell us about the statistical relationship between clinical depression and cancer?

Answer

The findings tell us that there is a statistical relationship between clinical depression and cancer.

The relationship could be in any one of the following four:

People that have clinical depression have a higher incidence of cancer.

People that have cancer have a higher incidence of clinical depression.

Spurious. That is, due to sampling variability.

The relationship could be due to another variable or variables.

Correlational designs

Question 4

If a correlational study finds a relationship between two variables, could you ever conclude that there is a causal relationship between the two variables?

Answer

No, in a correlational design you cannot conclude causality.

Scatter plots

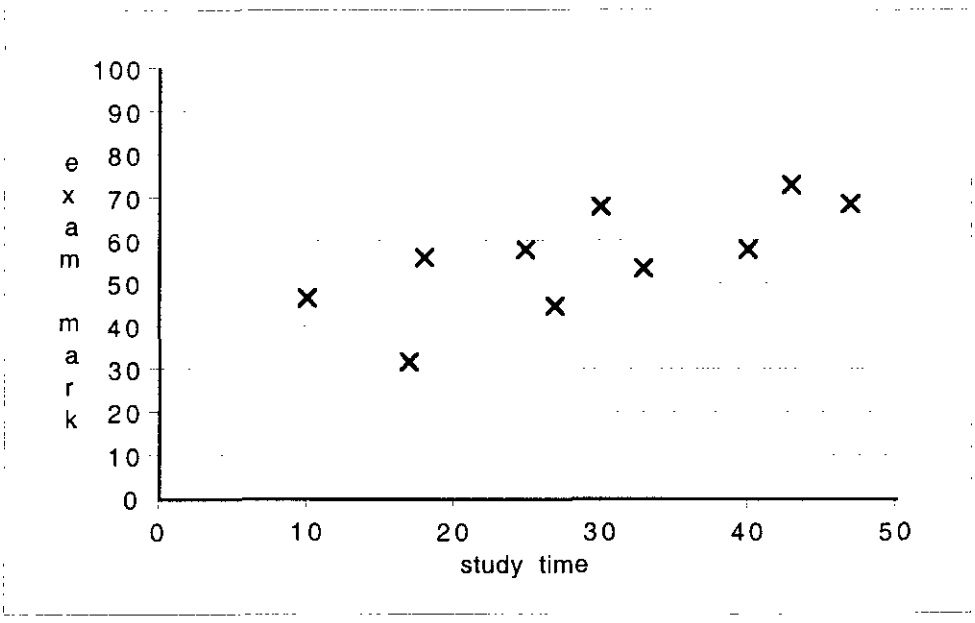
Question 5

The data in Table 1 gives findings from a study of ten first year university students showing how much time they spent studying (on average per week throughout the year) and their end of year examination marks (out of 100). Plot the data on the graph (Figure 1) to make a scatter plot.

Table 1 Students' study time and examination marks

Student	study time	exam. mark
1	40	58
2	43	73
3	18	56
4	10	47
5	25	58
6	33	54
7	27	45
8	17	32
9	30	68
10	47	69

Figure 1 Scatter plot of study time by exam. performance



(Plots not provided to students in booklet)

Answer

Figure 1. Scatter plot of study time by exam. performance (with plots)

Positive correlation

Question 6

What value does a perfect positive correlation coefficient take?

Answer

1, 1.0

Negative correlation

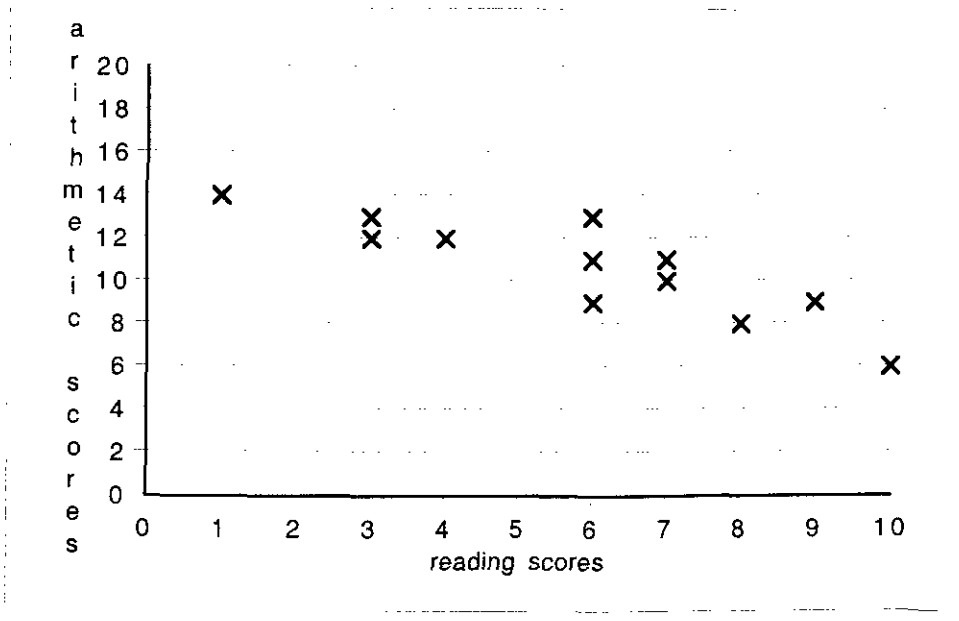
Question 7

In a study looking at the relationship between children's scores on a reading test and their scores on an arithmetic test, the data shown in Table 2 was obtained. Plot the data on the graph. (Figure 2).

Table 2 Children's arithmetic and reading scores

Child	reading score	arithmetic scores
1	9	9
2	7	10
3	8	8
4	10	6
5	7	11
6	6	9
7	3	12
8	4	12
9	6	13
10	6	11
11	3	13
12	1	14

Figure 2 Scatter plot of reading scores by arithmetic scores



(Plots not provided to students in booklet)

Answer

Figure 2. Scatter plot of reading scores by arithmetic scores (with plots)

Negative correlation

Question 8

What does the scatter plot show about the relationship between the two sets of scores?

Answer

There is a tendency for high scores on arithmetic to go with low scores on reading and high scores on reading to go with low scores on arithmetic as shown by the downward slope of the dots. This is known as a negative correlation.

Zero correlation

Question 9

Which of the following are most likely to result in a high positive correlation and which are not likely to be correlated at all?

Students' height and weight.

Girls' shoe size and scores on a reading test.

Students' self-ratings of ambitiousness and students' heights.

The number of theatre tickets sold and the number of customers in the audience.

Answer(s)

In general, there is usually a high positive correlation between students' height and students' weight. (See, for example, Jennings, Amabile & Ross, 1982).

One is very unlikely to find a correlation between girls' shoe size and scores on a reading test.

One is very unlikely to find a high correlation between students' self-ratings of ambitiousness and students' heights. (See Jennings et al, 1982, where it is reported that in one survey a correlation coefficient of .01 was found).

There should be a very strong correlation between these two variables because the number of tickets sold should correspond to the number of customers!

Zero correlation

Question 10

What is a likely correlation coefficient that you might obtain that would indicate no relationship between two variables. (For example, between students' self-ratings of ambitiousness and students' heights).

Answer

Zero or something very near zero.

The strength of a correlation

Question 11

Which of the following five correlation coefficients represent the greatest amount of correlation?

0.5, -0.8, 0.2, -0.4, 0.

Answer

-0.8

The strength of a correlation

Question 12

List the 5 correlation coefficients in order from those that indicate little or no correlation to that which indicates the greatest amount of correlation.

0.5, -0.8, 0.2, -0.4, 0.

____ no correlation

____

____

____

____ greatest amount of correlation

Answer

0 no correlation

0.2

-0.4

0.5

-0.8 greatest amount of correlation

The null hypothesis

Question 13

For a study that was to look at the relationship between students' examination performance (measured by scores on a particular examination) and students' performance on course work (measured by marks for an assignment), state the null hypothesis.

Answer

There is no relationship between students' examination performance and students' performance on course work. That is, the null hypothesis states that the population correlation coefficient is 0.

Parametric test: Pearson product moment correlation coefficient

Question 14

A psychologist was interested in the relationship between people's memory for shapes and their spelling ability, so she set up a study in which two tests were given to ten subjects. (Let's simply suppose that these two tests do in fact measure memory for shapes and spelling ability). The following Table 3 shows the scores that were obtained from the memory test for shapes and the test for spelling ability. State the null hypothesis and work out the value of the Pearson correlation coefficient, r. Use the formula for Pearson correlation coefficient provided.

Table 3 Subjects' memory test and spelling test scores

Subject	memory test	spelling test
1	6	3
2	7	4
3	8	2
4	9	4
5	4	1
6	3	1
7	9	5
8	8	5
9	10	6
10	11	5

Formula for Pearson correlation coefficient, r

$$r = \frac{N\sum a \times b - \sum a \times \sum b}{\sqrt{\left[N\sum a^2 - \left(\sum a\right)^2\right]\left[\left(N\sum b^2 - \left(\sum b\right)^2\right)\right]}}$$

where N = number of subjects

Answer

The null hypothesis states that there is no relationship between the two tests.

Student must calculate r by following an appropriate procedure.

Parametric test: the significance of a correlation coefficient

Question 15

Perform a two-tailed test to see if there is a significant relationship between the memory and spelling tests ($p < 0.05$). Here, use Table K provided.

Answer

For a two-tailed test ($p < 0.05$, d.f. = 8), the correlation coefficient value in Table K is 0.6319. For a correlation coefficient to be significant, it has to be equal of larger than this value in Table K. The calculated value 0.86 is therefore significant.

Parametric test: the significance of a correlation coefficient

Question 16

You have decided whether the calculated value of *r* is significant or not significant. What does your decision mean?

Answer

The calculated value of *r* is significant. The null hypothesis can be rejected. This provides support to the idea that there is a relationship between people's memory for shapes and their spelling ability.

Interpretation of data

A real life concern that is often given media coverage is the effects of television on children's and teenagers' social behaviour. Does watching violence on television encourage aggression? Much media debate surrounds this issue, but it is an important one in this day and age: it has been estimated that the average child in the USA, by the age of sixteen, will have spent more time watching television than being in school, and will have seen 13. 000 killings on television (Smith and Cowie, 1988). Psychologists have attempted to find out about the possible link that might exist between television violence and aggression. Let's take a study as an example.

A group of researchers interviewed the parents of children who were 9 years of age (184 boys, 175 girls) to see what they favourite television programmes were. From this, the

researchers constructed a measure of exposure to television violence. The children themselves were asked to rate the other children in their class for aggressiveness.

The researchers found that the correlation between these two measures was 0.21 for boys, but only 0.02 for girls. As shown in Figure 3 provided, the correlation for the boys was significant ($p < 0.01$).

Question 17i

What are likely explanations for these findings?

Figure 3 Correlations between the amount of television violence viewed at age 9 and peer-rated aggression at age 9 (184 boys and 175 girls)

Boys

TV violence
at age 9

0.21**

aggression
at age 9

** $p < 0.01$

Girls

TV violence
at age 9

0.02

aggression
at age 9

Answer

The significant correlation found for the boys could mean either than viewing television violence caused aggression, or that aggressive boys liked watching violent television programmes. Alternatively, other factors could be responsible for this correlation. For example, parental discord in the home could led a child both to watch violent television programmes and also be aggressive himself. (Other possible social influences could be stated: low income, low socio-economic class, parenting style, etc.) Explanations of gender differences in the findings might also be noted.

One might also point out methodological problems with the study. For example, the two measures are based on what the parents said about what their children's favourite television programmes and other children's ratings of a child's behaviour. These measures could be described as inaccurate or unreliable.

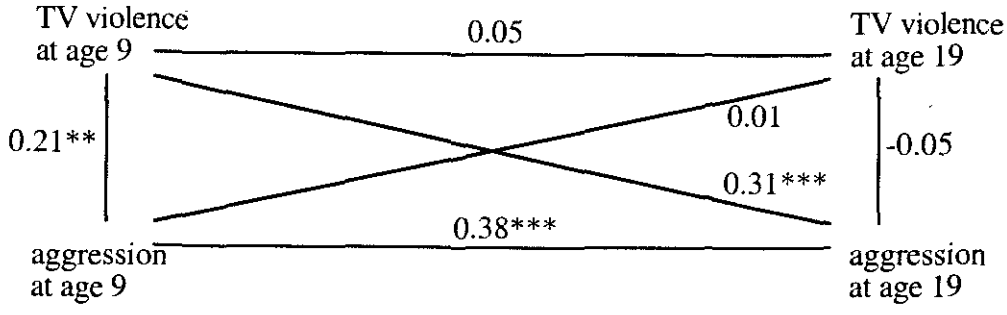
Question 17ii

Ten years later when the children were teenagers (19 years old), the same measures were taken. The correlations between the same two measures at this time *and* the correlations between the two different time periods are shown for both males and females in Figure 4 provided.

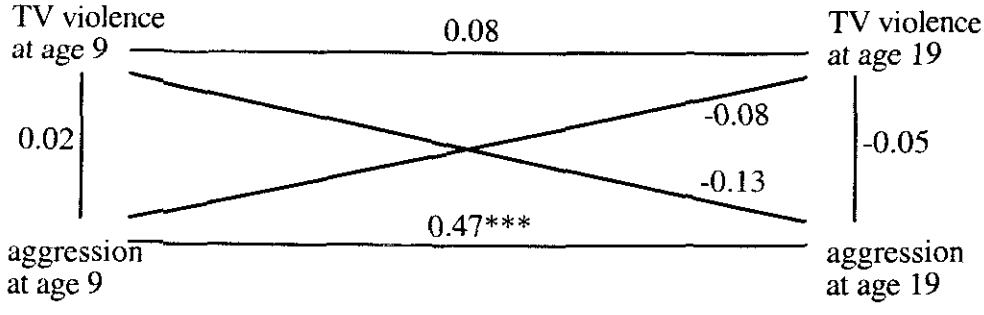
What do these findings suggest?

Figure 4 Correlations between the amount of television violence viewed at ages 9 and 19 and peer-rated aggression at ages 9 and 19 (184 boys and 175 girls)

BOYS



GIRLS



**p<0.01 ***p<0.001

Answer

The findings show that watching a lot of violent television at age 9 is significantly correlated (0.31) with peer-rated aggression at age 19. But peer-rated aggression at age 9 is not correlated (0.01) with watching violent television at age 19. This provides support for the idea that violent television leads to or encourages aggression rather than vice versa. Some other factor or factors might also be responsible for the associations. (See Smith and Cowie, 1988).

Figure 1 adapted from Hinton (1995, p. 255).

Figure 2 adapted from Green and d'Oliveira (1982, p. 136).

Figure 3 adapted from Smith and Cowie (1988, figure 4.4, p. 112. From Lefkowitz et al, 1977).

Figure 4 adapted from Smith and Cowie (1988, figure 4.4, p. 112. From Lefkowitz et al, 1977).

Formula for Pearson product moment correlation coefficient adapted from Green and d'Oliveira (1982, p. 143).

Question 5 adapted from Hinton (1995, pp. 254 - 255).

Question 7 and question 8 adapted from Greene and d'Oliveira (1982, pp. 135 - 136).

Question 9 adapted from The Open University (1990, p. 6 and p. 22) and Jennings et al, (1982, p. 218).

Question 11 adapted from Greene and d'Oliveira (1982, p. 137).

Question 12 adapted from The Open University (1990, p. 7).

Question 14, question 15 and question 16 adapted from Greene and d'Oliveira (1982, p.

144).

Question 17(i) and question 17(ii) adapted from Smith and Cowie (1988, pp. 111 - 113). Based on Lefkowitz et al (1977). Cited and described in Smith and Cowie (1988).

Table 1 adapted from Hinton (1995, p. 255).

Table 2 adapted from Greene and d'Oliveira (1982, p. 135).

Table 3 adapted and devised from Greene and d'Oliveira (1982, 144).

Appendix C

A sample of Lingo scripts

Scripts for the first and second prototypes of *Link*

Example C.1

```
on mouseUp
  play movie "TVActivities2"
end
```

Script of the button 'TV violence activities'. When this button is selected by a user a movie is played. In this case, this means that the user is provided with the introductory screen of the TV violence study.

Example C.2

```
on mouseUp
  go to frame "F1"
end
```

Script of the button 'Done' that is provided for activity 1. When this button is selected by a user they are provided with feedback to activity 1.

Scripts for the final version of *Link*

Example C.3

```
on startMovie
  global gUserLog
end
```

This script defines the variable for the user log.

Example C.4

```
on mouseUp
  global gUserLog
  set gUserLog = gUserLog & RETURN & "Selects Health events and goes to outline of health events"
  play movie "Outline C"
end
```

Script of the button 'Health events'. When this button is selected by a user, it invokes the screen that provides the outline of the Health events study. For the user log, it is recorded that the user has selected the button and is therefore presented with the screen that outlines the study.

Example C.5

```
on mouseUp
  global gUserLog, gWriteObject
  -- Create instance for writing to user file.
  put FileIO(mNew,"write","User file") into gWriteObject
  set gUserLog = gUserLog & RETURN & "Quit at" & RETURN & the long date & RETURN & the short time
  -- Method mWriteString writes the contents of variable gUserLog to the file
  gWriteObject(mWriteString,gUserLog)
  -- Dispose of the instance.
  gWriteObject(mDispose)
  quit
end
```

Script of the button 'Quit'. A user can select this button to quit the application. When the user quits the user log is written to a text file.

Example C.6

```
on mouseUp
  global gUserLog
  if the hilite of member "one" = TRUE and the hilite of member "two" = TRUE and the hilite of member "three" = TRUE and the hilite of member "four" = TRUE then
    set gUserLog = gUserLog & RETURN & "Selected 1234 and given feedback 1234 cohen"
```

```
play movie "Feed1234 C"
else if the hilite of member "one" = TRUE and the hilite of member "three" = TRUE and
the hilite of member "four" = TRUE then
set gUserLog = gUserLog & RETURN & "Selected 134 and given feedback 134
cohen"
play movie "Feed134 C"
else if the hilite of member "two" = TRUE and the hilite of member "three" = TRUE and
the hilite of member "four" = TRUE then
set gUserLog = gUserLog & RETURN & "Selected 234 and given feedback 234
cohen"
play movie "Feed234 C"
else if the hilite of member "one" = TRUE and the hilite of member "two" = TRUE then
set gUserLog = gUserLog & RETURN & "Selected 12 and given feedback 12 cohen"
play movie "Feed12 C"
else if the hilite of member "one" = TRUE and the hilite of member "three" = TRUE
then
set gUserLog = gUserLog & RETURN & "Selected 13 and given feedback 13 cohen"
play movie "Feed13 C"
else if the hilite of member "one" = TRUE and the hilite of member "four" = TRUE then
set gUserLog = gUserLog & RETURN & "Selected 14 and given feedback 14 cohen"
play movie "Feed14 C"
else if the hilite of member "two" = TRUE and the hilite of member "three" = TRUE
then
set gUserLog = gUserLog & RETURN & "Selected 23 and given feedback 23 cohen"
play movie "Feed23 C"
else if the hilite of member "two" = TRUE and the hilite of member "four" = TRUE then
set gUserLog = gUserLog & RETURN & "Selected 24 and given feedback 24 cohen"
play movie "Feed24 C"
else if the hilite of member "three" = TRUE and the hilite of member "four" = TRUE
then
set gUserLog = gUserLog & RETURN & "Selected 34 and given feedback 34 cohen"
play movie "Feed34 C"
else if the hilite of member "one" = TRUE then
set gUserLog = gUserLog & RETURN & "Selected 1 and given feedback 1 cohen"
play movie "Feed1 C"
else if the hilite of member "two" = TRUE then
set gUserLog = gUserLog & RETURN & "Selected 2 and given feedback 2 cohen"
play movie "Feed2 C"
else if the hilite of member "three" = TRUE then
set gUserLog = gUserLog & RETURN & "Selected 3 and given feedback 3 cohen"
```

```
    play movie "Feed3 C"
  else if the hilite of member "four" = TRUE then
    set gUserLog = gUserLog & RETURN & "Selected 4 and given feedback 4 cohen"
    play movie "Feed4 C"
  else
    set gUserLog = gUserLog & RETURN & "None selected and given activity 1 alert
cohen"
    alert "Click to select the available options. Select a maximum of four options."
  end if
end
```

Script of the button 'OK' that is provided for activity 1. Having selected available options on activity 1, a user can select this button to receive relevant feedback. For example, if the user only selects option 1 on the activity, they are then provided with feedback that emphasises that the other three available options are also possible interpretations of a correlation.

Example C.7

```
on mouseUp
  global gUserLog
  set gUserLog = gUserLog & RETURN & "Scatter plot 0.60 cohen"
  play frame "060"
end
```

Script of an object for the correlation 0.60, which is displayed in a table of data that is provided with activity 2. If the object is selected by a user, a screen, which presents the scatter plot representing a correlation of 0.60 and associated feedback, is invoked. This is recorded for the user log.

Example C.8

```
on mouseUp
 global gUserLog
 if sprite 37 intersects 43 and sprite 38 intersects 44 and sprite 42 intersects 45 and sprite
41 intersects 46 and sprite 39 intersects 47 and sprite 40 intersects 48 then
   set gUserLog = gUserLog & RETURN & "Feedback A cohen"
   go to frame "Fa"
 else if sprite 39 intersects 48 then
   set gUserLog = gUserLog & RETURN & "Feedback B cohen"
   go to frame "Fb"
 else
   set gUserLog = gUserLog & RETURN & "Feedback C cohen"
   go to frame "Fc"
 end if
end
```

Script of the button 'OK' that is provided with activity 3. This script checks to see how a user has arranged the correlation coefficients. If the correlations are arranged correctly, the user is provided with feedback (a). If the user has positioned the positive correlation in the strongest relationship position, they are provided with feedback (b). If the user has provided an undefined arrangement they are provided with feedback (c). (See appendix F).

Example C.9

```
on mouseUp
 global gUserLog
  repeat with gLoop1 = 48 down to 43
   repeat with gLoop2 = 42 down to 37
    if sprite (gLoop1) intersects sprite (gLoop2) then set gUserLog = gUserLog & "
sprite " & gLoop1 & " intersects " & " sprite " & gLoop2
   end repeat
 end repeat
 set gUserLog = gUserLog & RETURN & "Goes to outline cohen"
 play movie "Activity twoC"
end
```

Script of the button 'Activity 2' that is provided for navigational purposes on the screen of activity 3. When this button is selected a user is provided with the screen that presents activity 2 for the Health events study. The script records the user's arrangement of correlation coefficients on activity 3 for the user log. Part of the script is therefore used for all of the navigational buttons provided with activity 3.

Appendix D

Evaluation questionnaire

1. What do you think was the best thing about the program?

_____

_____

_____

_____

2. What do you think was the worst thing about the program?

_____

_____

_____

_____

3. What do you think needs changing in the program?

_____

_____

_____

_____

4. How easy did you find the tasks?

_____

_____

_____

_____

5. What did you think of activity 1 in the program?

_____

_____

_____

_____

6. What did you think of activity 2 in the program?

_____

_____

_____

_____

7. What did you think of activity 3 in the program?

_____

_____

_____

_____

8. Is there anything else that you would like to mention?

_____

_____

_____

_____

(Questions adapted from Hix & Hartson, 1993, p. 309; Monk et al, 1993, p. 84).

Appendix E

Tests in correlation used for the formative evaluation

Test A

Name

Date

1. If decreases in the X variable are accompanied by decreases in the Y variable, then the correlation between X and Y is positive. True or False?

(a) True

(b) False

1. (i) Explain your answer

2. If a correlational study finds a relationship between two variables, could you ever conclude that there is a causal relationship between the two variables?

(a) No

(b) Yes

(c) Sometimes

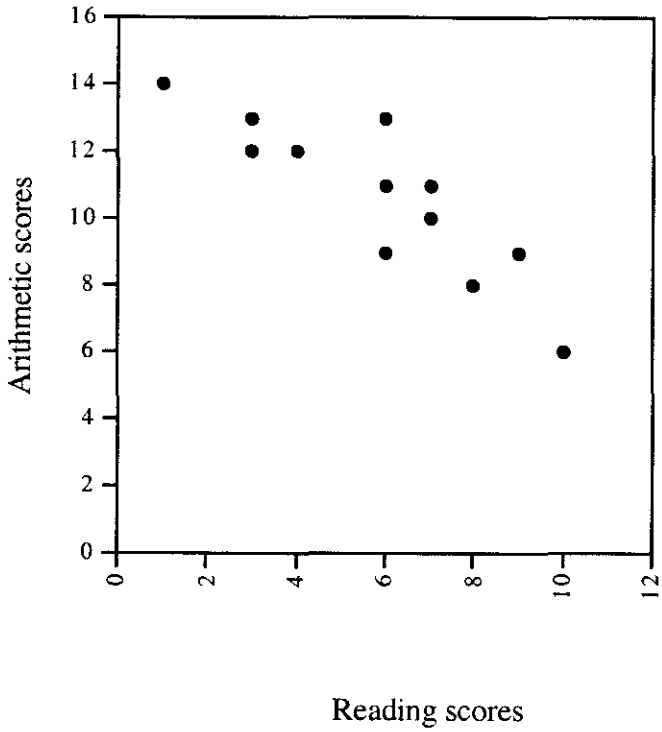
2. (i) Why?

3. In a study looking at the relationship between children's scores on a reading test and their scores on an arithmetic test, the data shown in the table were obtained. These data were plotted on a scatter plot. What does the scatter plot show about the relationship between the two sets of scores?

Children's arithmetic and reading scores

Child	reading score	arithmetic scores
1	9	9
2	7	10
3	8	8
4	10	6
5	7	11
6	6	9
7	3	12
8	4	12
9	6	13
10	6	11
11	3	13
12	1	14

Scatterplot of reading scores by arithmetic scores



4. Which correlation coefficient is stronger?

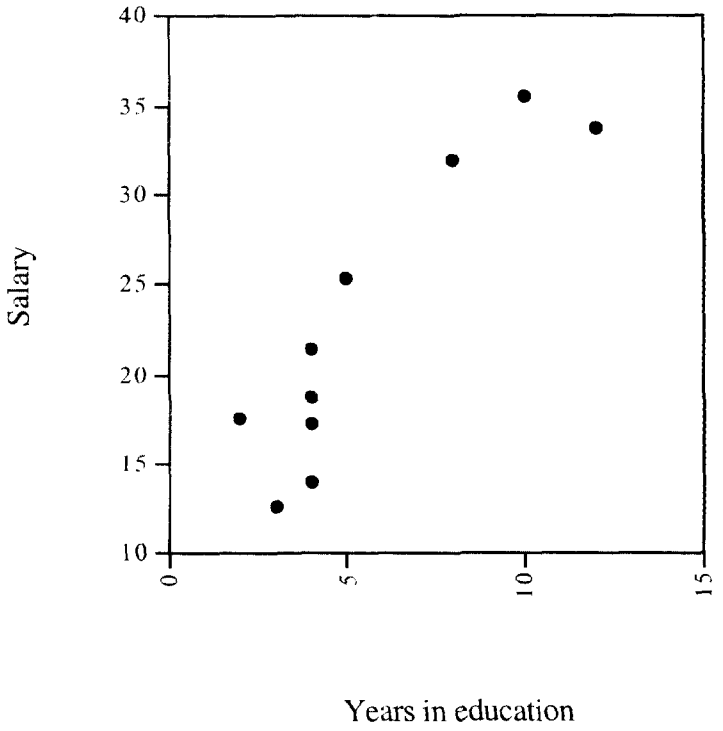
(a) 0.01

(b) 0.64

4. (i) Explain your answer.

5. What does the scatter plot show about the relationship between years spent in education and salary potential?

Scatterplot of salary by years in education



(a) That there is a positive relationship between years spent in education and salary potential

(b) That there is a negative relationship between years spent in education and salary potential

(c) That there is little or no relationship between years spent in education and salary potential

6. A group of researchers studying the relationship between creative thinking and intelligence administered different measures of creative thinking and intelligence to a sample of high school students. They obtained a correlation coefficient of 0.8 and concluded that high intelligence results in high scores on creative thinking. Is this conclusion warranted from the data?

(a) Yes

(b) No

6. (i) Explain.

7. What is a likely correlation coefficient that you might obtain that would indicate no relationship between two variables. (For example, between girls' shoe size and scores on a reading test).

7. (i) Explain your answer.

8. Which correlation coefficient is stronger?

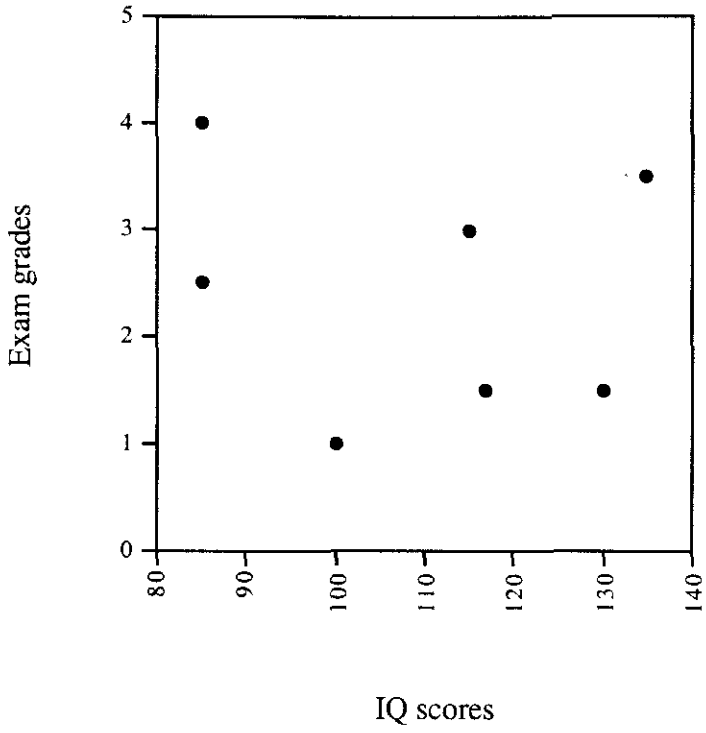
(a) -0.82

(b) 0.04

8. (i) Explain your answer.

9. What does the scatter plot show about the relationship between students' IQ scores and exam grades?

Scatterplot of students' IQ scores and exam grades



(a) That there is a positive relationship between students' IQ scores and exam grades

(b) That there is a negative relationship between students' IQ scores and exam grades

(c) That there is little or no relationship between students' IQ scores and exam grades

10. Which of these shows a correlation?

0.64

-0.83

(a) The first

(b) The second

(c) Both

(d) Neither

10. (i) Explain you choice.

11. Which correlation coefficient is stronger?

(a) 0.71

(b) -0.81

11. (i) Explain your answer.

12. When working on a psychology project together, two students obtained two correlation coefficients from the same data. Sue obtained a coefficient of -0.45 and Jane obtained a coefficient of 1.02. Which student is certainly wrong?

(a) Sue

(b) Jane

(c) Can't tell

12. (i) Why?

13. Which of these shows a correlation?

-0.84

0.02

(a) The first

(b) The second

(c) Both

(d) Neither

13. (i) Explain you choice.

14. Which of the following sets of correlations correctly shows the strongest to the weakest relationship?

Strongest weakest

(a) -0.91, 0.83, 0.65, 0.03

(b) 0.03, 0.83, 0.65, -0.91

(c) 0.83, 0.65, -0.91, 0.03

(d) 0.83, 0.65, 0.03, -0.91

14. (i) Explain your answer

Test B

Name

Date

1. A researcher obtained a correlation of 0.62 between the amount of time spent watching television and level of blood cholesterol. This means that there is a general tendency for people who watch less television also to have lower blood cholesterol. True or false?

(a) True

(b) False

1. (i) Explain your answer.

2. Professor Smith does an experiment and establishes that a correlation exists between variables A and B. Based on this correlation, she asserts that A is the cause of B. Is this assertion correct?

(a) No

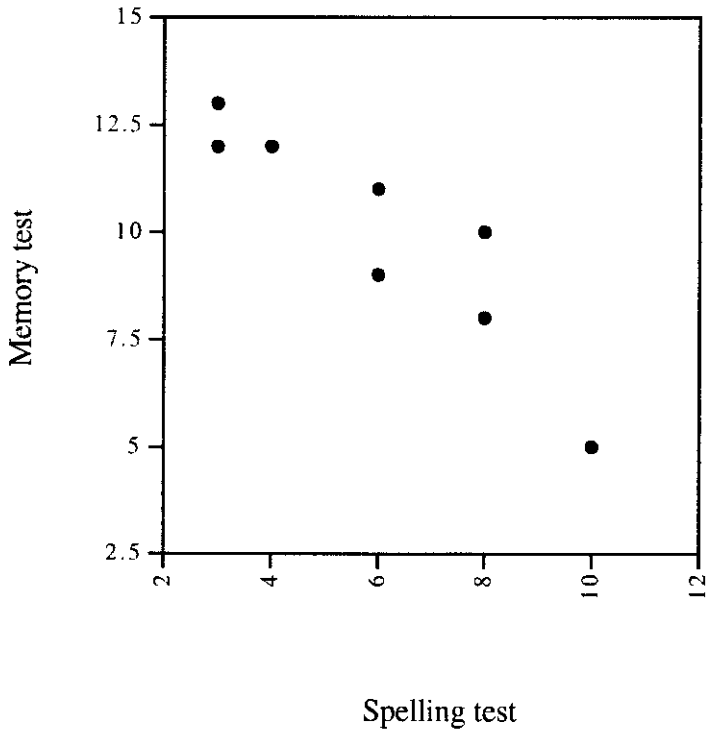
(b) Yes

2. (i) Explain.

3. In a study looking at the relationship between children's scores on a memory test and a spelling test, the data shown in the table were obtained. These data were plotted on a scatter plot. What does the scatter plot show about the relationship between the memory test scores and the spelling test scores?

Subject	Memory test	Spelling test
1	6	11
2	8	8
3	10	5
4	8	10
5	6	9
6	3	12
7	4	12
8	3	13

Scatterplot of memory and spelling test scores



4. Which correlation coefficient is stronger?

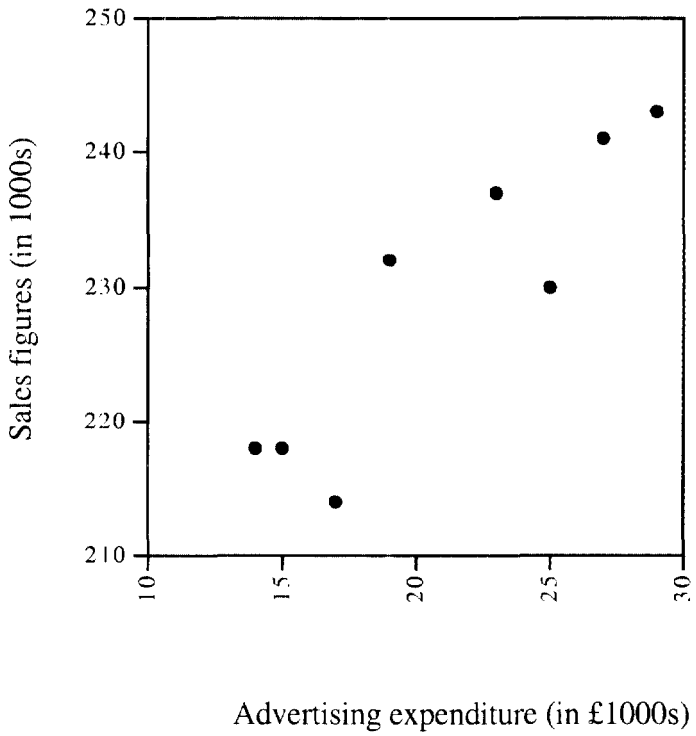
(a) 0.03

(b) 0.68

4. (i) Explain your answer.

5. What does the scatter plot show about the relationship between a company's advertising expenditure and sales figures?

Scatter plot of advertising expenditure and sales figures for the company



(a) That there is a positive relationship between advertising expenditure and sales figures for the company

(b) That there is a negative relationship between advertising expenditure and sales figures for the company

(c) That there is little or no relationship between advertising expenditure and sales figures for the company

6. Suppose there is a correlation of 0.87 between the length of time a person is in prison and the amount of aggression the person displays on a psychological inventory. This means that spending a longer amount of time in prison causes people to become more aggressive. True or false?

(a) False

(b) True

6. (i) Why?

7. What is a likely value of a correlation coefficient that would tell you that there is no relationship between two variables? (For example, between girls' shoe size and scores on a reading test).

7. (i) Explain your answer.

8. Which correlation coefficient is stronger?

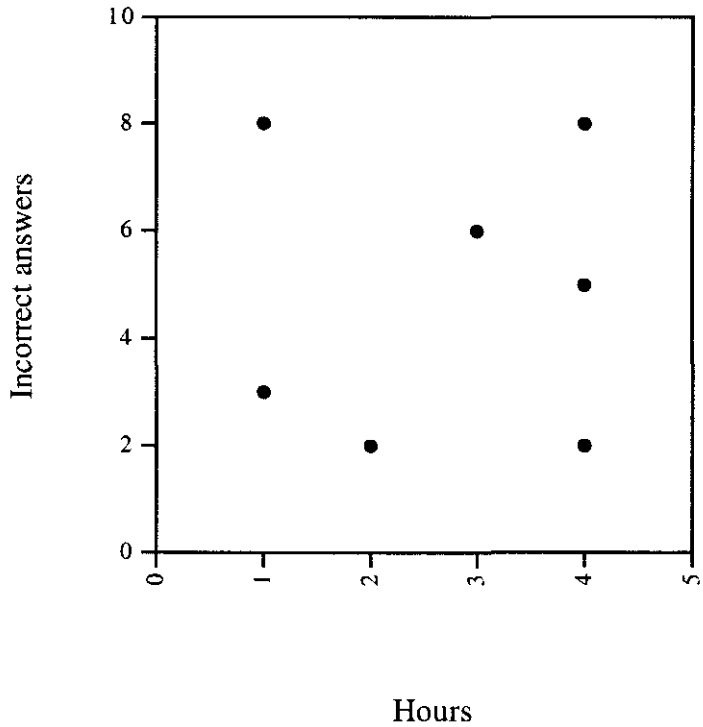
(a) -0.88

(b) 0.02

8. (i) Explain your answer.

9. What does the scatter plot show about the relationship between the number of hours that each student in a class had spent preparing for their exam and the number of incorrect answers on their exam papers.

Scatterplot of hours spent preparing and incorrect answers



(a) That there is a positive relationship between the number of hours that each student spent preparing for their exam and the number of incorrect answers on their exam papers

(b) That there is a negative relationship between the number of hours that each student spent preparing for their exam and the number of incorrect answers on their exam papers

(c) That there is little or no relationship between the number of hours that each student spent preparing for their exam and the number of incorrect answers on their exam papers

10. Which of these shows a correlation?

0.68

-0.85

(a) The first

(b) The second

(c) Both

(d) Neither

10. (i) Explain you choice.

11. Which correlation coefficient is stronger?

(a) 0.73

(b) -0.84

11. (i) Explain your answer.

12. When working on a their course work together, two students obtained two correlation coefficients from the same data. Jacqui obtained a coefficient of -0.57 and Jake obtained a coefficient of 1.08. Which student is certainly wrong?

(a) Jake

(b) Jacqui

(c) Can't tell

12. (i) Why?

13. Which of these shows a correlation?

-0.86

0.01

(a) The first

(b) The second

(c) Both

(d) Neither

13. (i) Explain you choice.

14. Which of the following sets of correlations correctly shows the weakest to the strongest relationship?

Weakest strongest

(a) -0.79, 0.56, 0.67, 0.04

(b) 0.04, -0.79, 0.56, 0.67

(c) 0.04, 0.56, 0.67, -0.79

(d) -0.79. 0.04, 0.56, 0.67

14. (i) Explain your answer.

Appendix F

Final design of *Link*

1. The purpose of the program

Link is designed to be used by psychology students who have covered the statistical topic of correlation. The program is a remedial program that is designed to address students' misconceptions concerning correlation.

Link provides an introductory screen that outlines the primary objectives of the program. *Link* also outlines two psychological studies:

- Memory for medical history (Cohen & Java, 1995).
- Infant engagement (Oates, 1998).

The program contains two sections, each of which provide:

- A screen providing a brief outline of the study. This includes a description of the variables.
- A table of data containing correlation coefficients.
- Three learner activities that use data from the study.

In addition, text in the program briefly outlines:

- That there are different correlation coefficients (e.g., Pearson or Spearman) and that the choice of correlation coefficient depends on the type of data collected in a study.
- That a correlation coefficient can be tested to see if it is statistically significant.

2. The human-computer interface

Link's human-computer interface provides:

- Navigational facilities. For example, a user is able to move from one activity to another activity without having to first invoke the introductory screen.
- An 'OK' button for each of the activities. When this is selected, a user is provided with appropriate feedback to the learner activity.
- When a user has finished a learner activity, text is provided that informs the user that they have completed the activity and that they can move on to complete another activity.

3. Generic design of activity 1 (causalistic)

- The basic design of this activity was maintained from the second prototype, but the question and options for this activity have been re-worded.
- This activity provides audio and text feedback.
- The feedback to this activity emphasises why causality cannot be inferred from a single correlation.
- This activity uses an additional example that is designed to address a causalistic conception of correlation. In the example, a correlation has been obtained between two variables, but it is clear that one variable cannot cause the other.
- The feedback is contingent on a user's response(s).

4. Generic design of activity 2 (unidirectional)

- The basic design of this activity was maintained from the second prototype.
- Students might not know that a correlation coefficient is represented by a scatter plot and the program makes this explicit.
- This activity provides graphical (scatter plot) and text feedback.
- If a user selects the correlation coefficient that represents the target scatter plot they are informed of this.
- When a user has selected a correlation coefficient, a scatter plot that represents the correlation is displayed. Feedback is also provided which:

  (i) states what kind of relationship is represented by the correlation coefficient selected by a user (e.g., negative correlation); and

  (ii) describes the relationship represented by the correlation coefficient selected.

5. Generic design of activity 3 (strength)

- The basic design of this activity was maintained from the second prototype.
- Feedback to this activity makes it clear that when assessing the strength of correlation coefficients, it is important to consider the size and the direction of a relationship.
- This activity provides text feedback and feedback in the form of an arrangement of correlations.
- The feedback is contingent on a user's response(s).

6. Introductory screen

In this package you will review your understanding of correlation. The aim of this package is to make sure that you have a clear idea about the different kinds of relationships that can be found between variables.

7. Memory for medical history (Health events study)

For this section of the program, the data set of a study conducted by Cohen and Java (1995) was used. By using the data set, correlation coefficients were computed and scatter plots were generated. Correlation coefficients were computed by using SPSS (version 6) and the application CA-Cricket Graph III was used to generate the scatter plots, which were then imported to the Director application.

Outline of the study

A study was conducted to investigate people's memory for health events and to look at measures of health status.

A sample of 104 people completed a health status questionnaire and kept health diaries for three months where they recorded the incidence, frequency and date of health events (e.g., illness, symptoms). Participants' memory for recorded health events was tested after the diary keeping period.

This study looked at the following measures:

Age. Participants' age.

Anxiety. Participants' levels of anxiety derived from their answers to questions on the health status questionnaire.

Correct dating. A score of the proportion of health events that were dated correctly by the participant. The dating of a health event was scored as correct if it was within two weeks of the date recorded in the diaries.

Depression. Participants' levels of depression derived from their answers to questions on the health status questionnaire.

Health events. The total number of health events recorded in the dairies.

Recalled events. The number of correctly recalled health events were scored as a proportion of the total number of events recorded in the diaries.

SRHS daily. Self ratings of health status that were recorded on a daily basis. Participants indicated this on a 4 point scale of very well-well-not very well-ill (where 1 = very well and 4 = ill). The mean derived from these daily ratings gave the measure called SRHS daily.

The correlations obtained in the study

	Age and depression	Anxiety and correct dating	Anxiety and depression
Correlations	-0.14	-0.21	0.60
	Depression and health events	Recalled events and health events	SRHS daily and health events
Correlations	0.28	-0.35	0.30

Click on the red text in the table (e.g., Age and depression) to find out about the measures in the study.

[Correlations]. When clicked the following feedback is provided:

> The table shows correlation coefficients (e.g., 0.60). A correlation coefficient provides a measure of the relationship between two variables.
> The correlation coefficients used in this program are the Pearson coefficient and the Spearman rank coefficient. The choice of coefficient depends on the type of data collected in a study.
> A correlation coefficient can be tested to see if it is significantly far from zero for a given sample size. For a sample size of say 40, the correlation 0.36 is statistically significant at $p < 0.05$.

Health events study. Activity 1

In the study a correlation of 0.60 was found to exist between participants' levels of depression and their levels of anxiety. What are the possible interpretations of this particular finding?

Select a maximum of four options.

1. That the participants' depression caused them to be anxious.

2. That the participants' anxiety caused them to be depressed.

3. That the correlation between depression and anxiety is spurious.

4. That another variable or variables could be responsible for the correlation.

Feedback provided is contingent on a user's selection.

If user selects 1 only. FW, F0, F2, F3, F4, F5.
If user selects 2 only. FW, F0, F1, F3, F4, F5.
If user selects 3 only. FW, F0, F1, F2, F4, F5.
If user selects 4 only FW, F0, F1, F2, F3, F5
If user selects 1 and 2. FW, F0, F3, F4, F5.
If user selects 1 and 3. FW, F0, F2, F4, F5.
If user selects 1 and 4. FW, F0, F2, F3, F5.
If user selects 2 and 3. FW, F0, F1, F4, F5.
If user selects 2 and 4. FW, F0, F1, F3, F5.
If user selects 3 and 4. FW, F0, F1, F2, F5.
If user selects 1, 2, 3 and 4. FC, F0, F5.

Feedback:

FC
Yes, all four options are possible.
FW
All four options are possible.
F0
From a correlation it is not possible to conclude that one of the variables such as depression has a direct causal affect on another variable such as level of anxiety. If a relationship is found to exist between two variables there are four possible interpretations.
F1.
It is possible that the participants' depression caused them to be anxious.
F2.
It is possible that the participants' anxiety caused them to be depressed.
F3.
It is possible that the correlation between participants' levels of depression and anxiety is spurious. This means that the finding was simply due to sampling variability or from sampling, for example, unusual behaviour.
F4
It is possible that another third variable could be responsible for the obtained correlation.
F5
Have a think about what kind of third variable could be responsible for the relationship that was found to exist between levels of depression and anxiety.

Brain size and IQ

One study found a positive correlation between a measure of brain size and IQ scores (Willerman et al, 1991). This does not of course mean that brain size is the cause of IQ, or vice versa. The correlation could be spurious, or other variables could be responsible for the correlation.

Willerman et al (1991) In Vivo Brain Size and Intelligence. *Intelligence*, 15, 223 - 228.

Health events study. Activity 2

	Age and depression	Anxiety and correct dating	Anxiety and depression
Correlations	-0.14	-0.21	0.60
	Depression and health events	Recalled events and health events	SRHS daily and health events
Correlations	0.28	-0.35	0.30

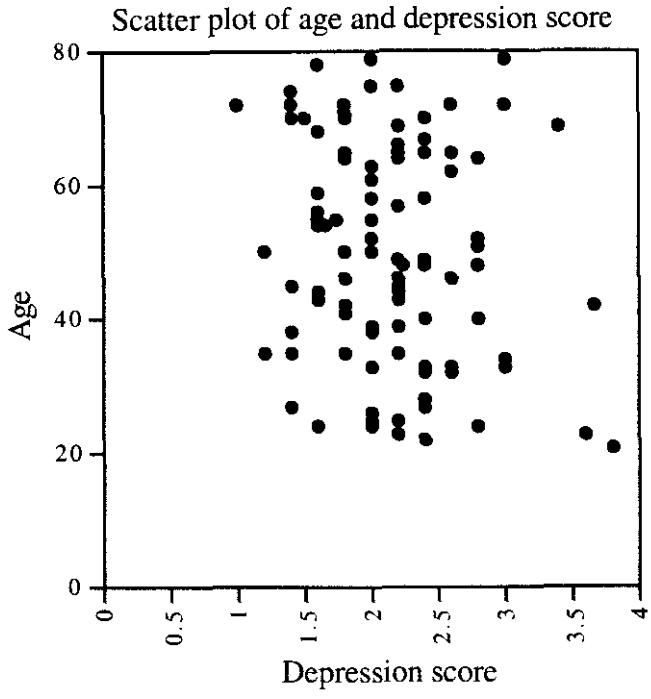
A correlation coefficient represents the data in any scatter plot by a single value.

Which correlation coefficient in the table represents the pattern on the scatter plot?

Click to select the correlation coefficient in the table.

Age and depression -0.14

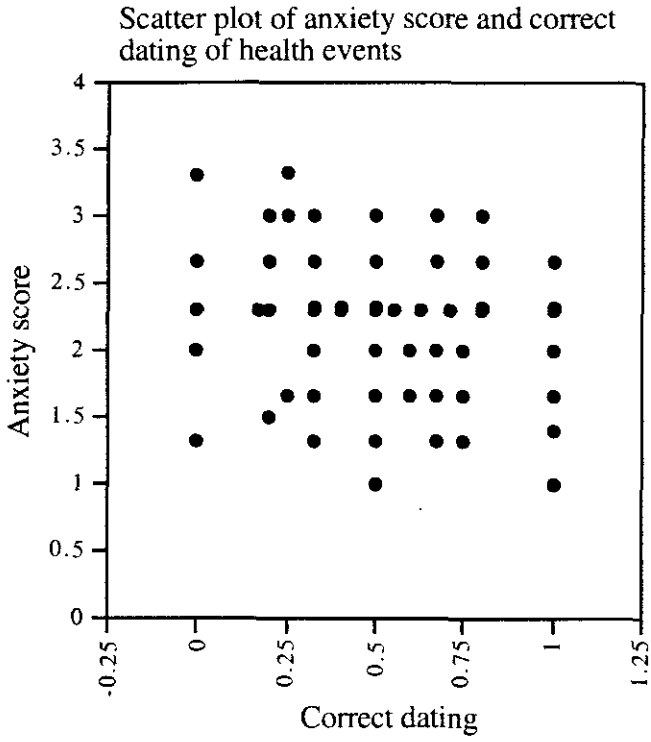
Feedback to -0.14 selection



-0.14 is a very weak negative correlation coefficient that is near to zero.
-0.14 indicates that there is a very weak correlation between the two variables.
When there is a very weak negative correlation between two variables, changes in one variable are not generally related to changes in the other variable.

Anxiety and correct dating -0.21

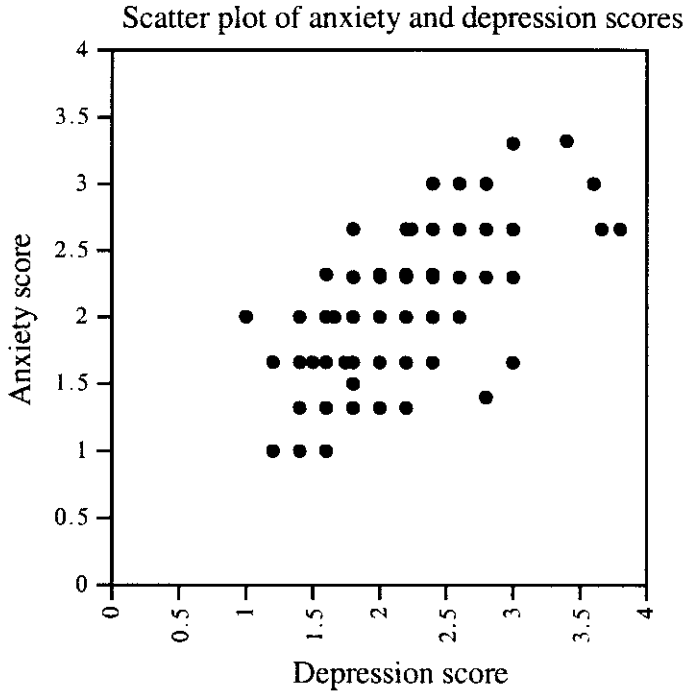
Feedback to -0.21 selection



-0.21 is a weak negative correlation coefficient.
-0.21 indicates that there is a weak correlation between the two variables.
When there is a weak negative correlation between two variables, changes in one variable are not generally related to changes in the other variable.

Anxiety and depression 0.60

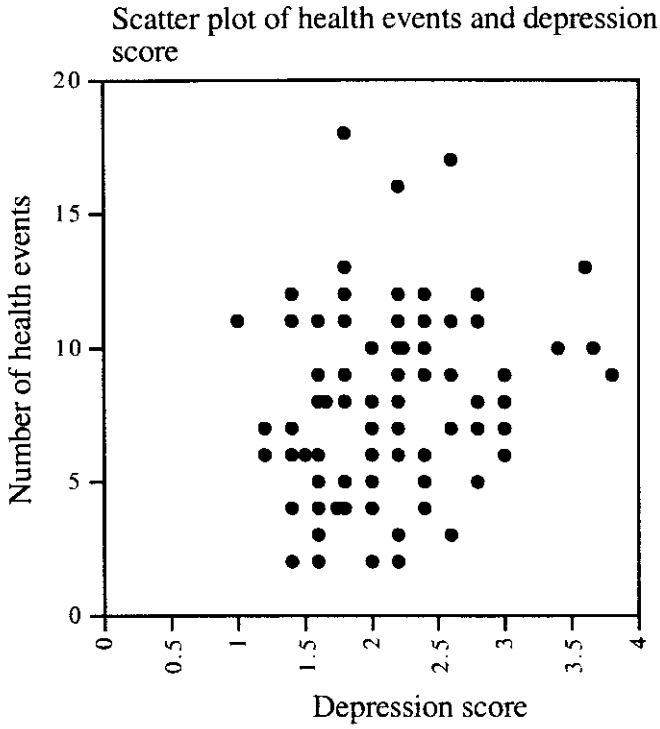
Feedback to 0.60 selection



0.60 is a strong positive correlation coefficient.
0.60 indicates that there is a strong correlation between the two variables.
A strong positive correlation indicates that increases in one variable are generally related to increases in another variable. This means that decreases in one variable are generally related to decreases in the other variable.

Depression and health events 0.28

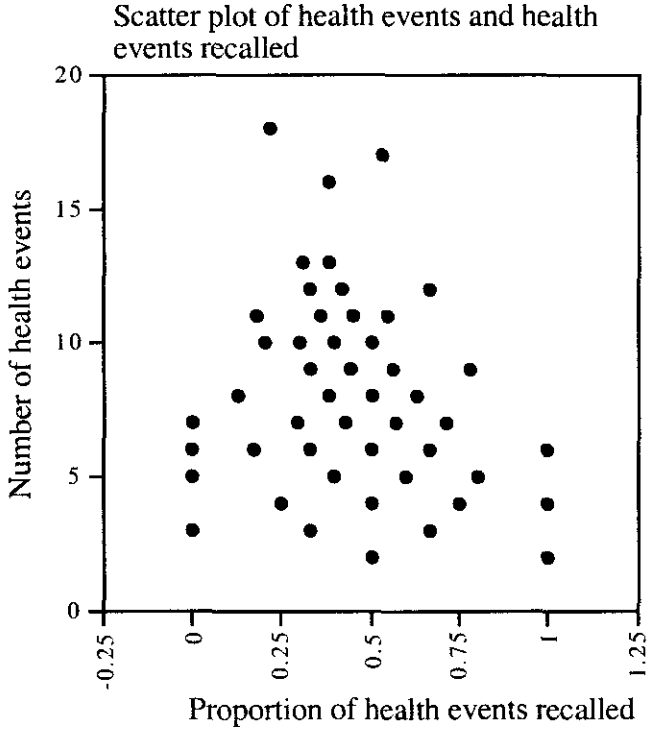
Feedback to 0.28 selection



0.28 is a weak positive correlation coefficient.
0.28 indicates that there is a weak correlation between the two variables.
When there is a weak positive correlation between two variables, changes in one variable are not generally related to changes in the other variable.

Recalled events and health events -0.35

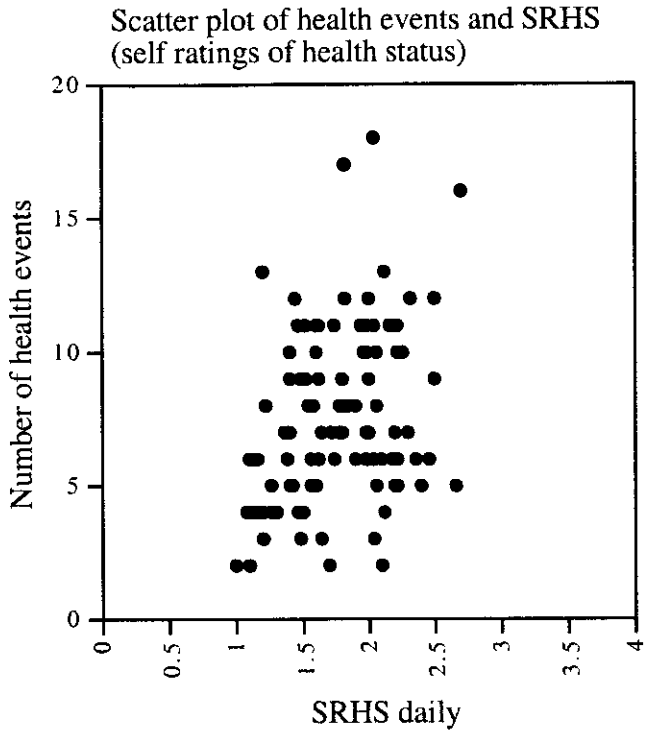
Feedback to -0.35 selection



-0.35 is a moderate negative correlation coefficient that represents the pattern on the scatter plot.
-0.35 indicates that there is a moderate correlation between the two variables.
A moderate negative correlation indicates that increases in one variable are generally related to decreases in the other variable, or that decreases in one variable are generally related to increases in the other variable.

SRHS daily and Health events 0.30

Feedback to 0.30 selection



0.30 is a weak positive correlation coefficient.
0.30 indicates that there is a weak correlation between the two variables.
When there is a weak positive correlation between two variables, changes in one variable are not generally related to changes in the other variable.

Health events study. Activity 3

Arrange the six correlations coefficients in the table in order from that which represents the weakest relationship between variables to that which indicates the strongest relationship between variables.

Click on the correlation coefficients in the table to select them. To arrange them, you can then drag the correlation coefficients from the table.

(For this activity, replace two correlation coefficients in the table).

	Age and depression	Example	Anxiety and depression
Correlation	-0.14	0.18	0.60
	Example	Recalled events and health events	SRHS daily and health events
Correlation	-0.65	-0.35	0.30

Correct ordering

-0.14 Weakest relationship
0.18
0.30
-0.35
0.60
-0.65 Strongest relationship

If correct ordering: Feedback (a)

Yes, a correlation coefficient that takes a value very near zero (e.g., -0.14) indicates that a very weak relationship exists between variables. The correlation -0.65 represents the strongest relationship between variables.
A correlation coefficient represents both the direction and magnitude of a relationship between variables.
The correlation coefficients have been arranged correctly.

If user positions 0.60 as the strongest relationship: Feedback (b)

No, the correlation -0.65 is stronger than the correlation 0.60. -0.65 indicates that a strong negative correlation or relationship exists between variables.
A correlation coefficient that takes a value very near zero (e.g., -0.14) indicates that a very weak relationship exists between variables. A positive or negative value can indicate that there is a very weak relationship between two variables.
A correlation coefficient represents both the direction and magnitude of a relationship between variables.
The correct arrangement of correlation coefficients is shown.

If any other positions: Feedback (c)

The correlation -0.65 represents the strongest relationship between variables. A correlation coefficient that takes a value very near zero (e.g., -0.14) indicates that a very weak relationship exists between variables. A positive or negative value can indicate that there is a very weak relationship between two variables.
A correlation coefficient represents both the direction and magnitude of a relationship between variables.
The correct arrangement of correlation coefficients is shown.

8. Infant engagement study

For this section of the program, the data set of a study conducted by Oates (1998) was used. Correlation coefficients were computed as described above. Scatter plots were generated and imported to the Director application as described above.

Outline of the study

A study was conducted with 43 mothers and their infants who were 2 months old. This study looked at maternal variables and infants' level of engagement in an experiment and in free-play with their mothers. For the experiment the infants were required to look at stimuli and their level of engagement was scored across a series of trials. Infants' level of engagement were also measured during a free-play episode with their mothers.

This study looked at the following measures:

EPDS. This was the mother's score on the Edinburgh Postnatal Depression Scale.

Free-play engagement. This was a measure of the infant's average level of engagement during free-play with the mother (e.g., eyes closed was scored as low engagement).

Education. The mother's education was scored on a scale of 1 to 7, where 1 indicated that the mother had no formal qualifications, and 7 indicated that the mother had post graduate qualifications.

Infant age. This was the infant's age in days.

Infant engagement. This was a measure of the infant's average level of engagement during the experiment.

Pregnancy reaction. This was a measure of the mother's reaction to finding out that they were pregnant. This was scored on a scale of 7 where 1 = negative disbelief and 7 = excitement.

Sole carer. This was the number of hours per week that the mother was the sole carer of her baby. This was measured by a daily record.

Table of data

	EPDS and free-play engagement	EPDS and pregnancy reaction	EPDS and sole carer
Correlations	-0.16	-0.36	0.11
	Infant age and infant engagement	Infant engagement and free-play	Education and sole carer
Correlations	0.56	0.60	0.05

Click on the red text in the table (e.g., EPDS and free-play engagement) to find out about the measures in the study.

[Correlations]. When clicked the following feedback is provided:

> The table shows correlation coefficients (e.g., 0.56). A correlation coefficient provides a measure of the relationship between two variables.
> The correlation coefficients used in this program are the Pearson coefficient and the Spearman rank coefficient. The choice of coefficient depends on the type of data collected in a study.
> A correlation coefficient can be tested to see if it is significantly far from zero for a given sample size. For a sample size of say 40, the correlation 0.36 is statistically significant at $p < 0.05$.

Infant engagement study. Activity 1

In the study a correlation of 0.56 was found to exist between infant age and infant levels of engagement. What are the possible interpretations of this particular finding?

Select a maximum of three options.

1. Infant age is causally related to levels of infant engagement.

2. The correlation between infant age and levels of infant engagement is spurious.

3. Another variable or variables could be responsible for the correlation between infant age and level of infant engagement in the experiment.

Feedback provided is contingent on a user's selection.

If user selects 1 only. FW, F0, F1, F2, F3, F4.
If user selects 2 only. FW, F0, F1, F3, F4.
If user selects 3 only FW, F0, F1, F2, F4.
If user selects 1 and 2. FW, F0, F1, F3, F4.
If user selects 1 and 3. FW, F0, F1, F2, F4.
If user selects 2 and 3. FW, F0, F1, F4.
If user selects 1, 2, and 3. FC, F0, F4.

Feedback:

FC
Yes, all three options are possible.
FW
All three options are possible.
F0
A single correlation coefficient provides a measure of the relationship between two variables. From a correlation it is not possible to conclude that one of the variables such as infant age has a direct causal affect on another variable such as a measure of infant engagement. If a relationship is found to exist between two variables there are four possible interpretations.
F1
It is possible that the infants' age caused them to be more or less engaged, but then it is possible that levels of infant engagement caused infants to become younger or older!
F2
It is possible that the correlation between infant age and infant levels of engagement is spurious. This means that the finding was simply due to sampling variability or from sampling, for example, unusual behaviour.
F3
It is possible that another third variable could be responsible for the obtained correlation.
F4
Have a think about what kind of third variable could be responsible for the relationship that was found to exist between infant age and infant levels of engagement.

Babies crawling and temperature

One study found a negative correlation between the average age at which babies crawl and average monthly temperature for the sixth month following birth (Benson, 1993). This does not of course mean that low monthly temperatures cause babies to crawl late. The correlation could be spurious, or other variables could be responsible for the correlation.

Benson, J. B. (1993) Season of Birth and Onset of Locomotion: Theoretical and Methodological Implications. *Infant Behavior and Development*, 16, 69 - 81

Infant engagement study. Activity 2

	EPDS and free-play engagement	EPDS and pregnancy reaction	EPDS and sole carer
Correlations	-0.16	-0.36	0.11
	Infant age and infant engagement	Infant engagement and free-play	Education and sole carer
Correlations	0.56	0.60	0.05

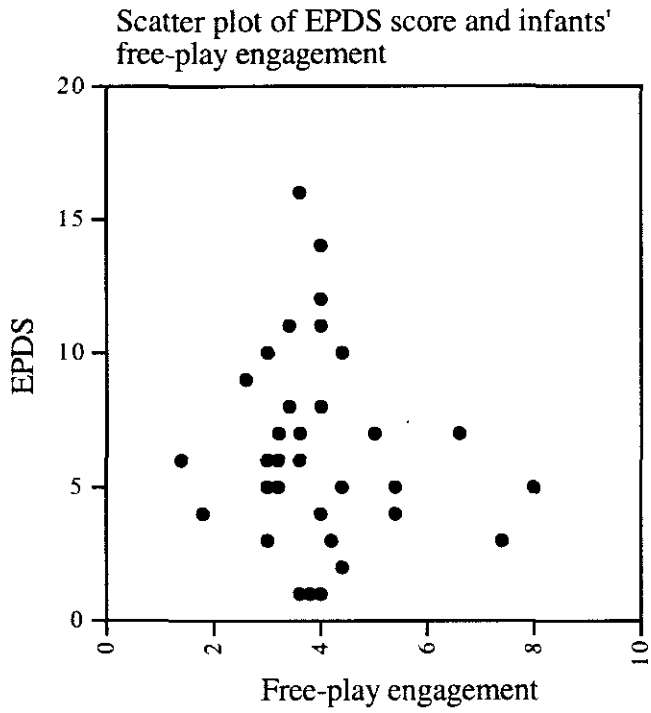
A correlation coefficient represents the data in any scatter plot by a single value.

Which correlation coefficient in the table represents the pattern on the scatter plot?

Click to select the correlation coefficient in the table.

EPDS and free-play engagement -0.16

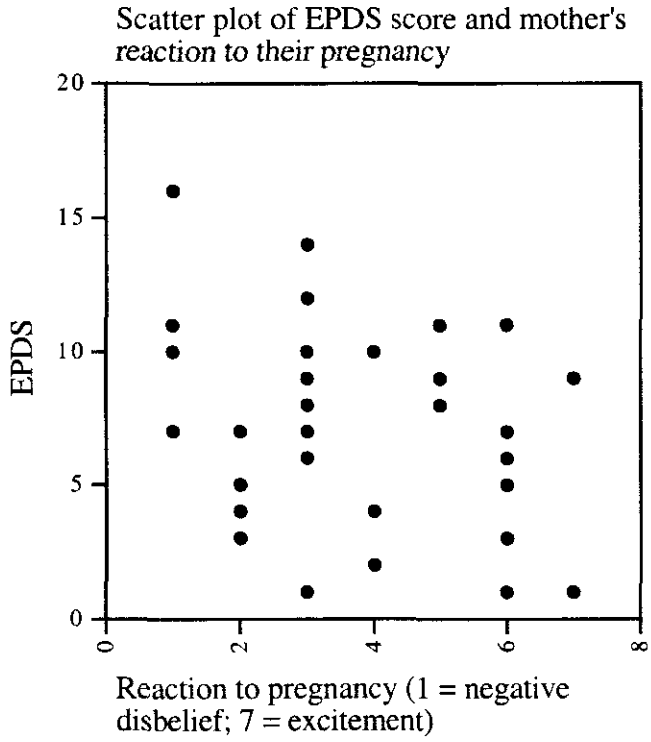
Feedback to -0.16 selection



-0.16 is a very weak negative correlation coefficient that is near to zero.
-0.16 indicates that there is a very weak correlation between the two variables.
When there is a very weak negative correlation between two variables, changes in one variable are not generally related to changes in the other variable.

EPDS and reaction to pregnancy -0.36

Feedback to -0.36 selection

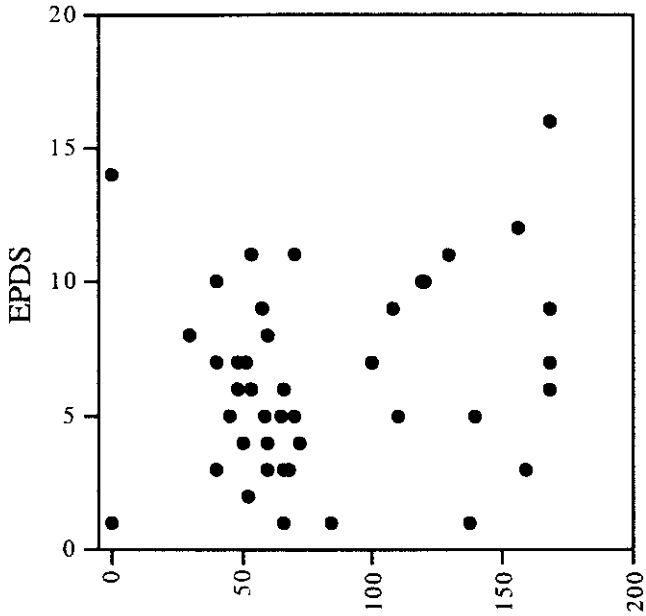


-0.36 is a moderate negative correlation coefficient that represents the pattern on the scatter plot.
-0.36 indicates that there is a moderate correlation between the two variables.
A moderate negative correlation indicates that increases in one variable are generally related to decreases in the other variable, or that decreases in one variable are generally related to increases in the other variable.

EPDS and sole carer 0.11

Feedback to 0.11 selection

Scatter plot of EPDS score and sole carer (hours per week)

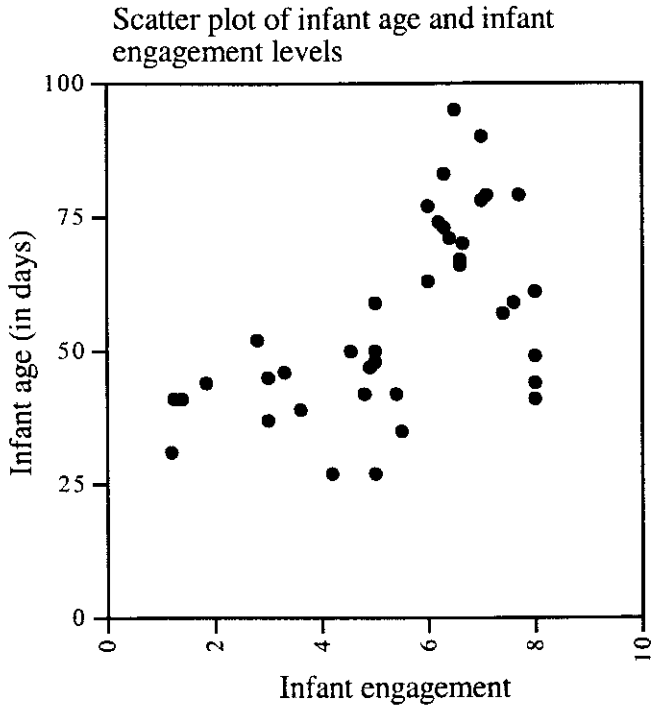


Number of hours per week that mother was sole carer of her baby

0.11 is a very weak positive correlation coefficient that is near to zero.
0.11 indicates that there is a very weak correlation between the two variables.
When there is a very weak positive correlation between two variables, changes in one variable are not generally related to changes in the other variable.

Infant age and infant engagement 0.56

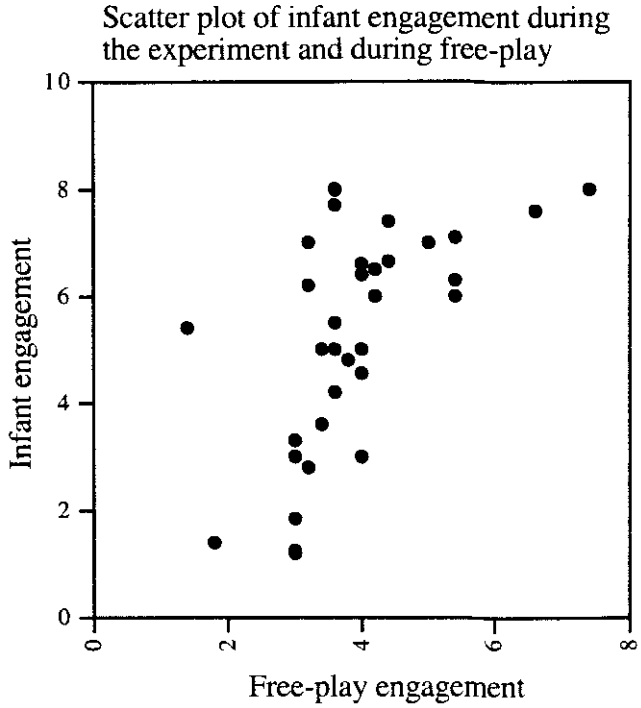
Feedback to 0.56 selection



0.56 is a strong positive correlation coefficient.
0.56 indicates that there is strong correlation between the two variables.
A strong positive correlation indicates that increases in one variable are generally related to increases in another variable. This means that decreases in one variable are generally related to decreases in the other variable.

Infant engagement and free-play engagement 0.60

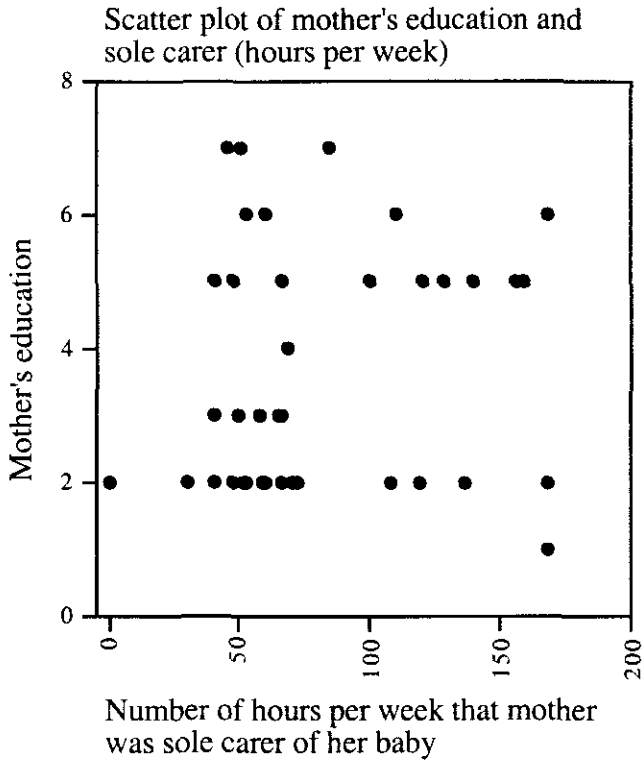
Feedback to 0.60 selection



0.60 is a strong positive correlation coefficient.
0.60 indicates that there is strong correlation between the two variables.
A strong positive correlation indicates that increases in one variable are generally related to increases in another variable. This means that decreases in one variable are generally related to decreases in the other variable.

Mother's education and sole carer 0.05

Feedback to 0.05 selection



0.05 is a very weak positive correlation coefficient that is near to zero.
0.05 indicates that there is a very weak correlation between the two variables.
When there is a very weak correlation between two variables, changes in one variable are not generally related to changes in the other variable.

Infant engagement study. Activity 3

Arrange the six correlations coefficients in the table in order from that which represents the weakest relationship between variables to that which indicates the strongest relationship between variables.

Click on the correlation coefficients in the table to select them. To arrange them, you can then drag the correlation coefficients from the table.

(For this activity, replace two correlation coefficients in the table).

	EPDS and free-play engagement	EPDS and pregnancy reaction	EPDS and sole carer
Correlations	-0.16	-0.36	0.11
	Example	Example	Education and sole carer
Correlations	0.32	0.26	0.05

Correct ordering

0.05 Weakest relationship
0.11
-0.16
0.26
0.32
-0.36 Strongest relationship

If correct ordering: Feedback (a)

Yes, a correlation coefficient that takes a value very near zero (e.g., 0.05) indicates that a very weak relationship exists between variables. The correlation -0.36 represents the strongest relationship between variables.
A correlation coefficient represents both the direction and magnitude of a relationship between variables.
The correct arrangement of correlation coefficients is shown

If user positions 0.32 as the strongest relationship: Feedback (b)

No, the correlation -0.36 is stronger than the correlation 0.32. -0.36 indicates that a moderate negative correlation or relationship exists between variables.
A correlation coefficient that takes a value very near zero (e.g., 0.05) indicates a very weak relationship exists between variables. A positive or negative value can indicate that there is a very weak relationship between two variables.
A correlation coefficient represents both the direction and magnitude of a relationship between variables.
The correct arrangement of correlation coefficients is shown.

If any other positions: Feedback (c)

The correlation -0.36 represents the strongest relationship between variables.
A correlation coefficient that takes a value very near zero (e.g., 0.05) indicates a very weak relationship exists between variables. A positive or negative value can indicate that there is a very weak relationship between two variables.
A correlation coefficient represents both the direction and magnitude of a relationship between variables.
The correct arrangement of correlation coefficients is shown.

References

Cohen, G. & Java, R. (1995) Memory for Medical History: Accuracy of Recall. *Applied Cognitive Psychology,* 9, 273 - 288.

Oates, J. (1998) Risk Factors in Infant Attrition and Low Engagement in Experiments and Free-Play. *Infant Behavior and Development*, 21, 4, 555 - 569.

Appendix G

A sample user log

(Participant S9 in summative evaluation study)

Thursday, March 19, 1998
2:34 pm
Selects Health events and goes to outline of health events
Goes to screen introduction
Selects Health events and goes to outline of health events
Selects age and depression cohen
Selects anxiety and correct dating cohen
Selects anxiety and depression cohen
Selects depression and health events cohen
Selects recalled events and health events cohen
Selects SRHS daily and health events cohen
Selects correlations cohen
Goes to activity 1 cohen
Selected 12 and given feedback 12 cohen
Goes to activity 2 cohen
Scatter plot -0.14 cohen
Scatter plot -0.21 cohen
Scatter plot 0.28 cohen
Scatter plot -0.35 cohen
Goes to activity 3 cohen
Feedback A cohen sprite 48 intersects sprite 40 sprite 47 intersects sprite 39 sprite 46 intersects sprite 41 sprite 45 intersects sprite 42 sprite 44 intersects sprite 38 sprite 43 intersects sprite 37
Goes to screen introduction
Selects infant engagement and goes to outline of infant engagement
Selects correlations oates
Selects EPDS and free play engagement oates
Selects EPDS and pregnancy reaction oates
Selects EPDS and sole carer oates
Selects infant age and infant engagement oates
Selects infant engagement and free play engagement oates
Selects education and sole carer oates
Goes to activity 1 oates
Selected 123 and given feedback 123 oates
Goes to activity 2 oates
Scatter plot 0.60
Scatter plot -0.36
Scatter plot -0.16
Scatter plot 0.56
Scatter plot 0.05
Scatter plot 0.11
Scatter plot -0.36
Goes to activity 3 oates
Given feedback A oates sprite 48 intersects sprite 38 sprite 47 intersects sprite 40 sprite 46 intersects sprite 41 sprite 45 intersects sprite 37 sprite 44 intersects sprite 39 sprite 43 intersects sprite 42
Quit at Thursday, March 19, 1998 2:48 pm

Notes

The date and times when the user starts and quits the program are recorded.

'Selected 12 and given feedback 12 cohen' means that the user has selected options 1 and 2 on activity 1 and has received feedback for this.

'Scatter plot -0.14 cohen' means that the user has selected the correlation -0.14 on activity 2 and the scatter plot displaying this relationship was shown on the screen.

'Feedback A cohen sprite 48 intersects sprite 40 sprite 47 intersects sprite 39 sprite 46 intersects sprite 41 sprite 45 intersects sprite 42 sprite 44 intersects sprite 38 sprite 43 intersects sprite 37' means that the user was provided with feedback A. This part of the log also specifies how the correlation coefficients were arranged from the weakest to the strongest relationship. On activity 3, each position is specified by a particular sprite, which is an instance of a cast member in a program, and each of the six correlation coefficients are also specified by particular sprites. Accordingly, it can be determined how the correlations are arranged. For example, 'sprite 48 intersects sprite 40' can be read as '-0.65 was placed in the strongest relationship position.'

Appendix H

Applications

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

ConStatS: Software for Conceptualizing Statistics (1997) Tufts University Curricular Software Studio.

Introduction to Research Design and Statistics. Demonstration version (1995) British Psychological Society. By Sandy MacRae.

SPSS (1989 - 1995) SPSS Inc., Version 6.1.

Statistics for the Terrified (1995) Stephen Morris, Jill Szuscikiewicz & Mark Preston. ITTI. St Bartholomew's Hospital Medical College.

Statistics Tutor: Tutorial and Computational Software for the Behavioral Sciences. Joseph D. Allen and David J. Pittenger. New York: John Wiley & Sons, Inc.

StatView (1992 - 1993), Abacus Concepts, Inc. Version 4. 5.

STEPS (1996) STEPS psychology module: *Predicting Dyslexia?* 26 Feb. 1996. Designed by P. Holmes, A. Simpson, E. Stillman & P. Derlien, and implemented by P. Derlien of the University of Sheffield. Produced under the auspices of the STEPS Consortium, funded by the Teaching and Learning Technology Programme of the UK Higher Education Funding Councils.

The Secrets of Psychology - Associative Learning. By N. W. Bond, School of Psychology, The Flinders. University of South Australia.

Understanding Statistics (1996) Royal Statistical Society. Developed by staff from the Centre for Statistical Education and the University of Sheffield (1990).