

https://dspace.org/

Computer-aided evaluation of television instruction in a tertiary-level introductory statistics course

Young, Derrick Aubrey

2014-07-25
http://hdl.handle.net/10539/15023

Downloaded from a DSpace repository. https://dspace.org/

COMPUTER-AIDED EVALUATION OF TELEVISION INSTRUCTION IN A TERTIARY-LEVEL INTRODUCTORY STATISTICS COURSE

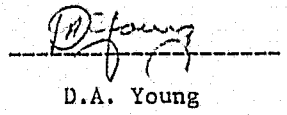
Derrick Aubrey Young

A Dissertation Submitted to the Faculty of Education, University of the Witwatersrand, Johannesburg for the Degree of Master of Education

Johannesburg 1979

Declaration

I hereby declare that this is my own work and that it has not been submitted to any other University.



D.A. Young

Abstract

This study investigated the effectiveness of televised lectures in teaching an introductory statistics course to first year commerce undergraduates. Certain constraints imposed themselves on this introductory course which dictated many of the teaching conditions. Some of these conditions were that the lecture situation formed the major teaching component in the course and that these lectures were administered to large numbers of students (groups of between 70 and 120).

The primary aim of the investigation was to determine whether or not the effectiveness of these lectures (when televised) could be assessed. A secondary aim was to determine whether or not (once the effectiveness of the lectures had been assessed) the areas in the lectures which had been identified as ineffective, could be analysed and changed, so as to become effective.

In order to conduct this investigation the method employed demanded that three preliminary areas be fully expounded. The first was that the meaning of 'instructional effectiveness' be fully defined in terms of the prevailing conditions. In order to achieve this a criterion referenced approach to instruction was adapted to the television lectures. Secondly, the 'type' of television lecture had to be disclosed. The televised lectures were traditional in that they were similar to the 'live' lectures except for a few novel innovations. The study did not however set out to compare the effectiveness of the 'live' and 'T.V.' instruction - a point which is fully discussed in the second and sixth chapters. The third preliminary area was the means employed to collect the data which was needed for the evaluation of the lectures. A recently developed educational computer system was used for this purpose and a full description of this novel system is given in this study.

The experimentation was based on two important premises. First, the evaluative means which were used to assess the lectures had to be valid. In other words they had to measure what they were supposed to

measure. The validation procedure adopted is therefore fully discussed. Secondly the variables had to be identified and controlled when improvements were attempted so as to ensure that the only variable which was allowed to change was the instruction. This procedure is always a difficult one and is fully discussed in chapter six and chapter eight.

The results of this investigation indicated that it was both possible to assess and improve the effectiveness of a televised lecture within the prevailing conditions. However this investigation is seen only as a preliminary study into an area which requires scientifically based analyses and conclusions in order to achieve both effective and efficient instruction in this teaching area. Therefore there is much which this study did not do and several criticisms are made in the final chapter.

To Bugs, who has been so patient.

Acknowledgements

Many thanks to Dr. David Daniels of the Chemistry Department for his guidance and assistance and to Mrs. Ellen van der Westhuizen for her diligent and co-operative manner in typing the manuscript.

Thanks are also expressed to colleagues in the Department of Applied Mathematics and the Department of Education for their assistance and co-operation, and to the staff of the Central Television Service for the production of the video tapes.

Contents Page

1. INTRODUCTION 1
1.1 An Orientation 1
1.2 The Structure 3
1.2.1 The Background 3
1.2.2 The Experiment 3
1.3 The Appendices 4

PART I: THE BACKGROUND

2. TEACHING WITH TELEVISION: A BRIEF REVIEW 7
2.1 The Development of Educational Television 7
2.1.1 Early History 7
2.1.2 Progress and Development 8
2.1.3 Various Uses 10
2.2 The Effectiveness of Television Instruction .. 11
2.2.1 Television and Face-to-Face Instruction 11
2.2.2 Experimental Inadequacies 13
2.2.3 Some Conclusions 15
2.3 A Case Study 16
2.3.1 The Design 16
2.3.2 The Results 19
2.3.3 The Interpretation 19

3. CRITERION-REFERENCED INSTRUCTION 21
3.1 The Systems Approach to Education 22
3.1.1 General Outline 22
3.1.2 Some Criticisms 24
3.2 A Criterion-Referenced Approach 21
3.2.1 The Ends 26
3.2.2 The Means 27
3.2.3 Evaluation Consideration 31
3.3 The Criterion-Referenced Approach to Television Instruction 36
3.3.1 A Delimitation 36
3.3.2 The Objectives of the Television Instruction 38

4. THE STRUCTURE OF THE TELEVISION PROGRAMMES 40
4.1 The Development of the Television Programmes ... 40
4.1.1 Initial Philosophy 40
4.1.2 The Optimum Length 43
4.1.3 Coloured Backgrounds 45
4.2 Some Cognitive Considerations 46
4.3 The Means of Evaluation 48
4.3.1 Intermittent Questions 48
4.3.2 The Value of Responding 50
4.3.3 The Primary Intention 51

5. A COMPUTER BASED TEACHING SYSTEM 55
5.1 The Structure of Components 56
5.1.1 The Mini-Computer and Disc Unit 58
5.1.2 The Instructor Console and Printer/Reader 58
5.1.3 The Student Panels 60
5.2 The Modes of Operation 64
5.2.1 The Social Mode 64
5.2.2 The Individual Mode 65
5.2.3 Results and Analyses 66
5.3 The Evaluation of the Television Programmes 66
5.3.1 The Lesson Programme 67
5.3.2 The Reports 69
5.3.3 Immediate Feedback 72

PART II: THE EXPERIMENT

6. THE DESIGN OF THE EXPERIMENT 76
6.1 The Purpose of the Experiment 77
6.1.1 Some Prevailing Conditions 77
6.1.2 The Model to be Tested 78
6.1.3 The Hypotheses 81
6.2 The Identification and Control of the Variables . 82
6.2.1 The Criteria for Evaluation 82
6.2.2 The Instructor Effect 83
6.2.3 The Response Delay 83
6.2.4 The Sampling Procedure 84

6.3 The Student Samples 84
6.3.1 The Measurement Variables 84
6.3.2 Pre-test Results 87

7. VALIDATION OF THE EVALUATIVE MEANS 92
7.1 Validation Procedures 92
7.1.1 Reliability and Validity 92
7.1.2 A Priori and A Posteriori Approaches ... 93
7.1.3 A Priori Validation of Criterion-Referenced Tests 94
7.1.4 The Validation Technique 96
7.2 The Matching of Questions and Objectives 97
7.2.1 The General Objectives or Aims 97
7.2.2 Specific Objectives 99
7.2.3 Matching Process 100
7.3 Results 100
7.3.1 The Validity Ratings 100
7.3.2 Discussion 103

8 THE RESULTS AND ANALYSES 105
8.1 The Preliminary Results 105
8.1.1 Attainment of the Acceptance Level 105
8.1.2 Some Hypotheses Tested 108
8.2 Analysis and Changes in Instruction 109
8.2.1 The Basis of the Analysis 109
8.2.2 The Method 112
8.2.3 A Summary 113
8.3 Final Results 117
8.3.1 The Question Scores 119
8.3.2 The Differences in Scores 125
8.3.3 The Tukey Test 132
8.3.4 Mann-Whitney U Test 137

9. THE CONCLUSION 146
9.1 The Instructional Intent 146
9.2 The Evaluation 148
9.3 The Improvement 149

APPENDIX I: The Questions and Objectives of the Television Lectures 152
APPENDIX II: Analyses and Instructional Changes 167
APPENDIX III: Transcripts of the Television Lectures 201
APPENDIX IV: Lesson Programmes and Computer Reports 273
APPENDIX V: The Pre-Test 280

Bibliography 285

Figures

		Page
Figure 1:	A Systems Approach to Education	23
Figure 2:	A Criterion-Referenced Instructional Paradigm ...	32
Figure 3:	Criterion and Norm Referenced Tests	34
Figure 4:	The Structure of Components in the Cybercom System	57
Figure 5:	The Instructor Console and Printer/Reader	59
Figure 6:	The three Components of the Instructor Console .	59
Figure 7:	The Mode Control Buttons	61
Figure 8:	Student Panels	61
Figure 9:	The Student Panel	62
Figure 10:	The Mode Choices	65
Figure 11:	The Lesson Programme for Lecture 1 - M35	67
Figure 12:	Expected Distributions for Identical Groups	86
Figure 13:	Expected Distributions for Non-Identical Groups	87
Figure 14:	Distribution of Scores for Group 1 (57 Questions)	110
Figure 15:	Graph of the T.V. Lecture Questions for Group 1	118
Figure 16:	Distribution of Results for Group 1 (55 Questions)	122
Figure 17:	Distribution of Results for Group 2 (55 Questions)	123
Figure 18:	Distribution of Results for Group 3 (55 Questions)	124
Figure 19:	Distribution of Instruction and DIFF1	129
Figure 20:	Distribution of Instruction and DIFF2	130
Figure 21:	Distribution of Instruction and DIFF3	131
Figure 22:	Graph for Tukey Test on DIFF1	135
Figure 23:	Graph for Tukey Test on DIFF2	136
Figure 24:	Graph for Tukey Test on DIFF3	138

Tables

		Page
Table 1:	Groups subjected to different treatments and lecturers	17
Table 2:	A Question Report	70
Table 3:	One way analysis of variance of the pre-test results for Groups 1,2 and 3	90
Table 4:	Significance Tests for Three Samples	91
Table 5:	Summary of matchings recorded by five content specialists	102
Table 6:	Group 1 Results for Television Lectures 1-6	107
Table 7:	Categorisation of Questions which did not Attain the Acceptance Level	115
Table 8:	Summary of Instructional Changes	115
Table 9:	Relationship between Changed, Unchanged, Initial and Final Instruction	119
Table 10:	Results of 55 Questions for Groups 1,2 and 3 ..	121
Table 11:	Results for Changed and Unchanged Instruction	126
Table 12:	Differences for Changed and Unchanged Instruction	128
Table 13:	Means for DIFF1, DIFF2, and DIFF3	132
Table 14:	Ranking of DIFF1	140
Table 15:	Ranking of DIFF2	142
Table 16:	Ranking of DIFF3	144

1. INTRODUCTION

> 'There are a number of ways of thinking about instruction. Some of them are conducive to instructional improvement. Others are not.'
> (Popham 1973, p9)

1.1 An Orientation

This study involved itself with evaluating the effectiveness of television instruction. This statement covers a vast field and a delimitation of the actual field of study is necessary. The learners were first year university students who were reading for a bachelor degree in commerce. An introductory statistics course was a compulsory requirement of their curriculum. A large portion of the teaching of this course was done by lecturing to large classes, because of various constraints which will be outlined. The lecture was presented to the students via the medium of television.

The evaluation which took place in this study measured the effectiveness of the television instruction in achieving its stated goals during the lecturing period. The teaching or learning which took place outside of the lecture was not evaluated.

The orientation which was taken in this study was based on the belief that the effectiveness of television instruction could be authentically determined and the results of the evaluation could be used to improve the instructional effectiveness. The experimentation conducted in this study was directed at testing hypotheses which were formulated from this orientation.

Any work which deals with the evaluation of instruction must involve three important aspects of teaching. The first deals with the criteria which are to be used to evaluate the instruction. One of the most difficult parts of evaluation comes at the beginning of a programme with the need to establish clear objectives; although this demands precision on the part of the course planners, unless such objectives are defined and accepted, evaluation may well prove a worthless exercise (Stephens and Roderick, 1971). Thus the criteria which were used are first clearly established. The second aspect involves the means which are used in the instruction.

The method used in the television instruction was based to a large extent on traditional lecturing methods. Two factors were major influences in this decision. First, it was desirous to measure just how effective traditional lecturing methods were in this statistics course and secondly, if the lecturing model developed in this study was to gain general acceptance, then the less radical the change, the better.

The final aspect is the development which takes place once the instruction has been evaluated. If no action is taken once the evaluative data is at hand, then the whole exercise is rather pointless. The efforts which were made in an attempt to improve the television instruction are fully described and the comparison of the initial and final instruction forms a major area of the experimentation undertaken.

The motivation for this work initially arose from a directive which requested that the effectiveness of the teaching in this introductory statistics course be optimized. Similar types of directives are becoming more common in higher education than ever before. There is more concern that educators should be accountable for their performance, and that education systems should attain the targets which they have been set (Lomax, 1979 (1)).

1.2 The Structure

The body of this work is divided into two parts. Part one deals with all the background features which led up to the experimentation and part two deals with the experiment itself. Some concluding remarks are then made in the final chapter.

1.2.1 The Background

Part one comprises chapters two, three, four and five. Chapter two gives a brief review of educational television and some conclusions which have been reached regarding its effectiveness. A case study of a typical type of instructional television experiment concludes the chapter.

Chapter three deals with the systems approach to education. A criterion-referenced instructional model is propounded which is based on the systems approach. The implementation of this model to the television lectures is discussed.

Chapter four outlines the development of the television instruction in the statistics course up until the time of the experiment. Some cognitive connotations in the development of the instruction are considered and the means which were used to evaluate the television lectures are revealed.

Chapter five concludes the background picture by outlining a new computer based teaching system. The use made of this sytem in order to evaluate the instruction is then described.

1.2.2 The Experiment

The second part of the study comprises three chapters which cover the experimentation that was conducted. Chapter six discusses the design of the experiment which includes the sampling procedure used, the identification and control of the variables and the enunciation of the hypotheses which were tested.

Chapter seven outlines the procedure which was used to validate the questions; these questions were required to reveal the level which the students had attained in regard to behaviourally stated objectives, during the television lectures.

Chapter eight begins with the results obtained from the initial evaluation. The analyses of and the changes made to the defective instructional sequences are discussed. The results of the final evaluation are then stated and several statistical tests are performed which determined the validity of the null hypotheses.

A final chapter, chapter 9, concludes the study with some inferences which are drawn from the experimental results. Several appendices follow this chapter.

1.3 The Appendixes

Six television lectures, which covered a clearly defined topic in the statistics course, were used in the study. After the initial evaluation, changes were made to parts of the six lectures in an attempt to improve their effectiveness. There were therefore twelve television lectures in all and the relevant cassette video tapes form an apparatus addendum to this work.

Appendix I gives the specific behavioural objectives for each lecture, in addition to the questions which were used to determine the attainment of these objectives. Included in this appendix is the form which content-specialists used to validate the questions.

Appendix II supplies a detailed description of the analysis which took place for each instructional sequence which failed to reach the acceptance level. The instructional changes which were made are stated in addition to the initial and final results achieved by the students for each question. Video tape reference points are also given so that these instruc-

tional sequences can be easily located and viewed.

Transcripts of the initial six lectures are given in appendix III. Any additions or changes which were made to the instruction when producing the final six television lectures are indicated in italics. This appendix should be used in conjunction with the viewing of the television instruction as the lectures are transcribed verbatim and the appendix will be difficult to follow if it is only read. Alternatively the appendix can be used to locate specific instructional sequences; video tape reference points are given at regular intervals.

Appendix IV supplies the lesson programmes which were used by the computer system in order to evaluate the television instruction. The nature of a lesson programme is expounded in chapter 5, where the computer system used in this study is described. Examples of the reports which this computer system generated are included after the lesson programmes in this appendix.

Finally, a copy of the pre-test is given in Appendix V. This pre-test was used to determine the level of attainment of certain mathematical skills which had been acquired by the students in the various samples, prior to the experimentation.

PART I :

THE BACKGROUND

2. TEACHING WITH TELEVISION: A BRIEF REVIEW

In the past twenty five years or so, television has been making an increasing appearance in education systems, both on a formal and informal basis, possibly due to the irreversible thrust of technology (Gordon, 1970). Although its basic form of presentation has essentially remained unchanged in its relatively short history, some major technological advances have occurred which have influenced and determined its use and developing use in education. Large amounts of time and money have been spent over the past two decades in order to evaluate its effectiveness as a teaching medium. It would seem appropriate then that some of the highlights in its development should be mentioned in an attempt to trace a path to its present position. This chapter will deal with some of the major uses and technical developments, followed by an examination of some of the evaluative research which has taken place. Finally a particular case study of the almost classical experiment will be presented: television versus traditional teaching.

2.1 The Development of Educational Television

2.1.1 Early History

The early history of television is probably well known: Zworykin's experiments during the 1920's; RCA's early interest in the technology of video and leadership by the BBC in Britain during the 1930's; and, finally, the first public television broadcasts of the opening of the

New York World's Fair in 1939 over NBC's experimental station, W2XBS (Gordon 1965).

It was in the 1950's that television made a significant appearance in education. Educational programmes were broadcast by the BBC in the UK and several educational television stations were founded in the USA. The growth of these educational stations continued into the 1960's until by 1965 there were 120 of these stations in the USA serving all levels of education (Carnegie 1967). Although television broadcasting for schools in the UK dates from 1957, little educational use was made of television in British universities at this time. However, rapid growth took place with the advent of the Brynmor-Jones report in 1965 on 'Audio-visual aids in higher scientific education' which recommended, amongst other things, the use of modern media and technology in teaching in higher education (Whitaker 1979).

The USA and the UK were the pace setters in the introduction of educational television, while other countries followed at a comparably slower pace. One country which progressed at a remarkably slower pace in this introduction was South Africa. Although technically capable of broadcasting a television service by the late 1950's the first television service was broadcast by the South African Broadcasting Corporation (S.A.B.C.) in January 1976. Educational television was introduced on a small scale in South Africa in the earlier part of this decade by some schools and universities. The University of the Witwatersrand was one of the pioneers in this direction when a closed-circuit television service was established on campus in 1970.

2.1.2 Progress and Development

In the middle 1960's, television equipment really suitable for studio production work at the closed-circuit level still did not exist.

Most media centres either had to buy the big, bulky and expensive broadcast equipment or else make some adaptation to industrial equipment. Video-tape recording had to be done on the rather unreliable helical-scan equipment or on the expensive broadcast quality machines (Whitaker 1979). The result was that most educational television programmes (apart from those broadcast by the large broadcasting corporations) tended to be as a live medium, which placed certain limitations on their use.

As the equipment costs of reliable machines became more reasonable, video-recording began to make a greater impact in educational television. In the early 1970's Japanese manufacturers started to design equipment specifically for non-broadcast use. There gradually appeared a wide range of cheaper purpose-built equipment for the small non-broadcast studio. Another development which came later was that of video-tape editing.

This meant that in the production of a programme, mistakes or problematic areas could be edited. It was no longer necessary to make an entire programme at one time, as the programme could be made in sections. These remarks are particularly revelant to this study as the experimentation could probably not have taken place without the editing facility in a non-broadcast standard studio.

Another important development in the middle 1970's was the advent of the cassette and cartridge video-tape recorders. This greatly facilitated the use of educational television programmes as not only did it allow for greater portability and easier use by technically unskilled teachers and students, but video recordings could now be made available in university libraries in such a way that students and staff could view them at a time when it was most convenient to them. (Whitaker 1979).

One of the latest technical developments in educational television is that of colour. Broadcast television in colour has been transmitted in many countries for quite some time but it is only recently that the cost of colour equipment has decreased sufficiently to make it a viable proposition for most media centres. The mass production of printed circuits has meant that the cost of electronic equipment has tended to fall in real terms. Extremely complex equipment can now be produced quite cheaply and the cost of a well-equipped colour studio is no longer such a prohibitive factor.

2.1.3 Various Uses

In a brief synopsis such as this on the development of educational television it would be an impossible task to describe the uses to which television has been put in the educational sphere since the 1950's. It would be fairer to point out that television has been used in most if not all of the teaching situations as we know them today.

The earlier held fear that television would be capable of replacing the teacher has proved to be largely unfounded. It has tended to be used more as an aid, supplement or convenience than a complete 'teaching entity'. Some of the uses to which it has been put are however highly questionable in terms of the necessity of its use. Has television been used in some areas because of its effectiveness or simply because it was there? The relentness demand which was felt during the 1960's for educational innovation led to the common tendency of equating innovation with improvement without rigorously evaluating the effectiveness of each new education programme or instructional procedure (Cohen, Rose and Trent, 1973).

2.2 The Effectiveness of Television Instruction

2.2.1 Television and Face-to-Face Instruction

The advent of educational television heralded some unique situations which had not been possible with educational film. At any one point in time a television programme could be shown in many areas simultaneously either via a closed circuit process (cables) or high frequency transmission. In addition to this a programme could be video-recorded and replayed at any time to a large number of groups or even to an individual student. The appearance of video cassette players made this process much easier. In addition the lecture theatre did not have to be darkened or a projector set up, as is the case with film.

Could educational television do the job as well as the 'live' teacher? If this was the case exciting possibilities existed: every-one could be taught by the 'expert'; teachers could be released from repetitive lessons and their time used more efficiently in other areas; the teacher could possibly be replaced!

It is no wonder then that, with these ideas being entertained, many studies have been conducted since the 1950's in which the effectiveness of television has been compared to that of the 'live' teacher. The overall results obtained could be described as disappointing. Far too many studies reported that there was no significant difference (N.S.D.) between the two. Gordon (1970) is severely critical of this fact, stating that it was like a new drug, N.S.D., in the 1960's for educational television researchers. He claims that between 1950 and 1970 organizations such as the Ford Foundation spent around a billion dollars on educational television research and they have very little to show for it. He quotes a report on the findings of educational television research: 'To our initial surprise and later disappoint-

ment, we found over and over again that there was no significant difference between television and conventional instruction.' (Gordon 1970: 203). A similar attitude can also be detected in other reports.

Stickell (1963) analyzed 250 studies which compared television instruction with face-to-face instruction and judged only 10 of them to be interpretable. In a summary of 100 studies of the effectiveness of television reviewed by Schramm (1964) it was found that 84 of the investigations reported no significant differences in achievement between televised and conventional instruction. Although many studies have shown no significant differences between the 'televised' and 'live' instruction, the overall result of all this research tends to show that television is no worse than face-to-face instruction and in several situations it is in fact more effective.

By the end of the 1960's the educational research had discovered that television could 'teach' as well (or as poorly) as the professor (Cohen and Trent, 1973). In a later review by Schramm (Chu and Schramm, 1967) it was found that sufficient experiments have shown the effectiveness of television to dispel any doubts that children and adults learn a great amount from instructional television, just as they do from any other experience that can be seen to be made relevant to them. Any earlier notions that television could do everything single-handed without the classroom teacher playing a vital role have largely disappeared and educational television is mostly accepted as an aid to supplement the teaching process. Many research studies comparing co-operative television teacher - classroom teacher with conventional instruction could be cited to show the resulting superiority of the 'team' approach (Moldstat, 1974).

2.2.2 Experimental Inadequacies

Although monumental amount of research has been conducted in order to compare instructional television with more conventional teaching methods, a disproportionately small number of meaningful or interpretable experiments have resulted. One possible reason for this is the lack of control of variables and the resulting poor experimental design in what is viewed by many as an area of extreme complexity and interaction. Many of these studies suffer from a variety of theoretical and methodological inadequacies (Lumsdaine, 1963).

One of the explanations for these results is that no one is quite sure exactly what it is that they are measuring (Gordon, 1970). Is one measuring powers of retention only or does T.V. have other values? And if so how does one set up an experiment that keeps other variables constant (like the influence of television at home)? In most of these studies no attempts were made to control certain variables such as type of instructional programme, type of conventional instruction, instructor effect, or appropriateness of random-selection procedures (Cohen, Rose and Trent, 1973).

As an illustration of the extreme complexity involved in a thorough and overall examination of the effects of instructional television, consider the case of 'Sesame Street', the American public television program directed towards pre-school children. Direct evaluation of the programme's goals indicated that children were making excellent progress toward the programme's objectives, such as the ability to recognize letters, the ability to count, the ability to group objects along similar dimensions etc. In a more recent study of 'Sesame Street' on Israeli kindergarten and school children, similar cognitive results were achieved once the children had mastered the mental skills necessary to overcome the modification caused by the novel formats

(Salomon, 1976). However no analysis of the overall effect of 'Sesame Street' was made.

Attempts to answer some of the following questions could and perhaps should have been made. Did the slick production tend to decrease children's tolerance for less exciting instructional programmes in any media? Did children develop increased confidence in their ability to learn? Did the passivity (the child as viewer not doer) have any discernible transfer to the child's manner of participation in ther instructional settings? Did children display reluctance to make responses in the presence of a personal, responsive figure, like the teacher, since they had enjoyed the anonymity of the friendly but distant television personalities? (Baker, 1973). These and other questions could be asked and when they are all considered the extreme difficulty involved in making a comprehensive evaluation of the effects of a widely distributed programme become apparent.

Studies conducted on a smaller more personalized scale can suffer from numerous other defects. If a teacher is aware that his or her instruction is to be compared with some other form of instruction by assessing student achievement at the end of the instruction, the teacher often tends to teach specifically what the test dictates. (Esson, 1972). There are many other variables which are often involved when evaluating educational television and their identification and control represents a major test in any such experiment. Experimental psychologists have never been very enthusiastic about evaluative experiments in which learning from programmes using the new media is compared with learning results based upon conventional instruction (Allen, 1971). Their primary objection centres on the inexactness in controlling variables associated with either the experimental (new media) approach or the control (conventional instruction) approach.

2.2.3 Some Conclusions

The vast majority of studies on the effectiveness of television instruction have reported no significant difference between this type of instruction and traditional instruction. This appears to be the situation involving most studies which compare a conventional teaching method and a method which makes use of some 'new media'. Most objectives may be attained through instruction presented by any of a great variety of media. A great number of studies have shown no significant difference between one medium and another in facilitating the attainment of a wide range of objectives (Levie and Dickie, 1973).

Many of these studies have suffered from a lack of precision in experimental design and control. The net result has been a famine of well structured guidelines on the implementation, use and analysis of educational television programmes. Some problems associated with the introduction and continued use of instructional television have been documented by Murphy and Gross (1966) which range from teacher antaganism to improper use of the medium.

This has left instructional television and indeed the whole field of instructional technology in somewhat of a dilemma. Two points of view are given by Moldstad (1974) where he describes some reports as stating that instructional technology is largely supplementary to the two primary media of instruction, the textbook and the teacher, and is something that the good teacher can quite easily manage without. If either the teacher or the textbook is eliminated, the whole education system would be transformed, but if all of technology is eliminated, education would go on without a missed lesson.

He then goes on to describe some of the distinct advantages that educational technology has, as is shown by media research over the past twenty years. The criticism of both points of view (for and

against 'new media') is that they are both based on 'soft data': utilization statistics, teacher testimonials or questionnaires, educated guesses, undocumented research claims, media attitude responses and student documentation of use and preferences

In today's world where there is a greater stress on issues such as cost- effectiveness it is important that more 'hard data' is produced in this research area. Although the concept of instructional television has gained prominence in many countries in the past two decades, there are only a few documented attempts to scientifically evaluate its success (Cohen, 1979)(1)).

Most of the points that have been made in this brief review of educational television are pertinent to this study. The area of evaluation must be clearly defined and there must be great precision in delimiting what it is about instructional television that is being investigated. If this is achieved the interpretation of the result can be made with greater certainty and clarity, in the knowledge that if the experiment is repeated under the same conditions with equal control over all the variables, the same result will be obtained.

To conclude this review a brief exposition will be given of an experiment undertaken to measure television instruction against traditional teaching. The experiment succumbed to many of the pitfalls described in the preceeding pages.

2.3 A Case Study

2.3.1 The Design

The Department of Applied Mathematics at the University of the Witwatersrand, Johannesburg, embarked on a project in 1978 whereby videotaped lectures were to replace 'live' lecturers in one of its

service courses. The year long service course is designed to give undergraduate students in the Faculty of Commerce an appreciation and working knowledge of certain mathematical and statistical concepts and applications which are of use in business, commerce and industry.

The project did not aim at removing the lecturer from the lecture theatre entirely. The format was one where a video recorded programme would be shown for only part of a 45 minute lecture period. The programmes were between 15 and 30 minutes in duration and the content covered the theoretical aspects of the course. As the lectures were pre-recorded and directed to different parts of the campus at any one time from a central transmission area, no control was exercised by the lecturer or the students over the replay. There was no 'live' interaction between the programme and the students. In the remaining time, the lecturer would apply the theoretical ideas developed in the television programme to the solution of practically orientated problems. Thus the format was not one of lecturer replacement but rather a team teaching approach between the 'televised' and 'live' lecturer. Some factors which influenced this choice of approach will be discussed in chapter 4.

This was the first time that a Department at the University had opted for television instruction to such an extent. The major part of every topic covered in the service course was to be taught by television. As a safe-guard against failing students who might claim that the instructional medium was the cause of their failure (amongst several other reasons) a study was to be conducted over the first trimester of the year to determine the effectiveness of the televised approach against the conventional (entirely 'live' lecturer) approach.

The evaluation was to be measured in terms of student achievement in tests set at the beginning and end of the trimester. The difference in the scores would serve to measure the gain of the individual student.

Six roughly equal groups (approximately 100 in each group) and two lecturers participated in the experiment. Each lecturer was responsible for two groups. The lecturer followed traditional teaching methods with one group and the 'television lecture - follow up' approach with the other. This formed a factori . part to the experiment. In other words not only could two teaching techniques be compared but also the differences between two lecturers and any interactions which could have occurred (Winer, 1971). This is more clearly illustrated by the following table:

		Treatment	
		A	B
Lecturer	1	Group α	Group β
	2	Group γ	Group δ

Table 1: Groups subjected to different treatments and lecturers.

The factorial design of the experiment thus allowed for measurements to be made between:

(a) treatment A and treatment B,

(b) lecturer 1 and lecturer 2,

(c) interaction between groups $(\alpha+\delta)$ and groups $(\beta+\gamma)$.

The remaining two groups were given lecturers and methods which could not be included in the experiment, but could provide an additional basis for error estimation and therefore would increase the precision of the main comparisons. The lecturers were aware that the experiment was taking place; the udents were not.

2.3.2 The Results

A detailed mathematical analysis of the results will not be given here (Young, 1978(1)). The final summary revealed that:

(a) there was no significant difference between the television instruction and the traditional instruction.

(b) there was no significant difference between the two lecturers.

(c) there was no significant interaction between the groups.

A closer examination shows that the largest differences occurred between the two lecturers but these were still so small that they were not at all significant. The 95% confidence limits were used.

2.3.3 The Interpretation

The result could almost have been predicted: no significant difference. Thus this experiment falls into the same category as so many similar studies have done in the past; there is no significant difference between television instruction and face-to-face instruction. But does this tell us that the television lectures were as good as (or as bad as) the 'live' lectures? A result such as this tells us very little that is worth knowing.

The Applied Mathematics Department was quite contented with the result as it shows that students did no worse when they were subjected to television lectures. However, as is typical with most of these experiments as outlined in section 2.2.2, there are so many variables which were not controlled that it makes this experiment very difficult to interpret.

No control was exercised over the amount of independent study which occurred during this period of time. Perhaps the television instruction was so bad that these groups were forced to do far more independent study than the groups who received the traditional teaching, and in so

doing managed to achieve results which showed no difference between the groups. Perhaps the television instruction was far superior and a reverse of the preceding situation occurred, yielding results which were still not significantly different. No measurement was made to determine to what extent the students relied on the lectures themselves. Perhaps they performed a minor role and no matter what happened in the lecture theatre, the results would not have differed. Perhaps the lecturers, being aware that the experiment was taking place, put far more preparation into their 'live' teaching and almost sabotaged the television lecture follow- up. These, and many more uncontrolled variables could be identified to make the result almost impossible to interpret in a meaningful way. This is probably a good example of Moldstad's 'soft data' referred to in section 2.2.3.

This case study has been cited as an illustration of many of the points raised in this chapter. It has also been reviewed because it acted as a forerunner to the experimentation in this work in that it revealed many inadequacies which can exist in an experimental design of this nature. There is a need for more precise, scientifically based research in the [illegible] of television instruction in mathematics and statistics. Television has been utilized in these subjects to a limited degree but there is an extreme dearth of available research pertaining to its effectiveness (Dessart and Frandsen, 1973).

3. Criterion Referenced Instruction

Lecturing to large groups is the most prevalent form of teaching in higher education and is usually the most difficult to reconcile with well established learning principles (University Teaching Methods Unit: 1978). The teaching of the course in statistics under investigation in this study consists by force of circumstances to a large degree of lecturing to large groups of students. The numbers vary from group to group between 60 and 120. The numbers in these groups might even grow substantially.

Thus having accepted that a large proportion of the teaching in this course will be conducted in the lecture theatre to large groups of students some questions in regard to the lecture itself must be asked. How effectively (and perhaps efficiently) are the teaching staff performing in these situations? How much teaching is actually done in a lecture and more importantly, how much does a student learn in a lecture of this nature?

As a starting point in order to answer some of these questions the teacher has to be more specific about what he hopes to achieve in the lecture. The introduction of television instruction to the lecture made the specifying of aims and objectives almost imperative. The six televised lectures which were constructed for this study adopted this approach, or more specifically a criterion-referenced approach.

This chapter will include a synopsis of criterion-referenced instruction and will then describe how this approach was implemented in the television lectures.

3.1 The Systems Approach to Education

3.1.1 General Outline

The systems approach was originally developed within the physical sciences because of the difficulty scientists had in conceptualising and expressing the characteristics of complex entities (Coles, 1979). Systems theory is thus concerned with the study of organised complexity and it can be argued that any education system or learning situation is such an entity. The basic argument rests on the proposition that a variety of different systems can be analysed and made amenable to planning by the use of a general theory (Birley, 1972).

There have been certain developments in education and psychology in the past two decades or more which have been dominant influences in the development of the systems approach. The first of these was the emergence of the concept of behavioural objectives. Tyler (1949) states that the real purpose of education is to bring about significant changes in the student's pat ern of behaviour. An attempt to produce a taxonomy of behavioural objectives in the cognitive domain was made by Bloom (1956) and a later taxonomy was attempted in the affective domain (Bloom, 1956; 1964). The same has been attempted for psychomotor objectives by Simpson (1967). Mager's (1962) contribution was to indicate how to define educational objectives. He felt that education without objectives was rather like setting out on a journey without knowing where you are going. The art of stating educational objectives in behavioural terms has been subject to much development, and two contributions

which are worthy of note are those of Ebel (1965) and Nedelsky (1965).

Another influence was the findings that were emerging from experimental psychologists in the area of learning. Most of this work was spearheaded by Skinner (1953) who claimed that efficient learning was possible and predictable if any piece of behaviour was appropriately rewarded; the main consideration being the scheduling of the reinforcement. One major application of this work can be seen in the development of programmed instruction. These developments, the so called technology of education, continued throughout the 1960's and when placed within a system theory, the result is a feedback-type systems model of education: objectives, design of learning, evaluation and improvement (Rowntree, 1974). The system expounded by Rowntree is illustrated in figure 1.

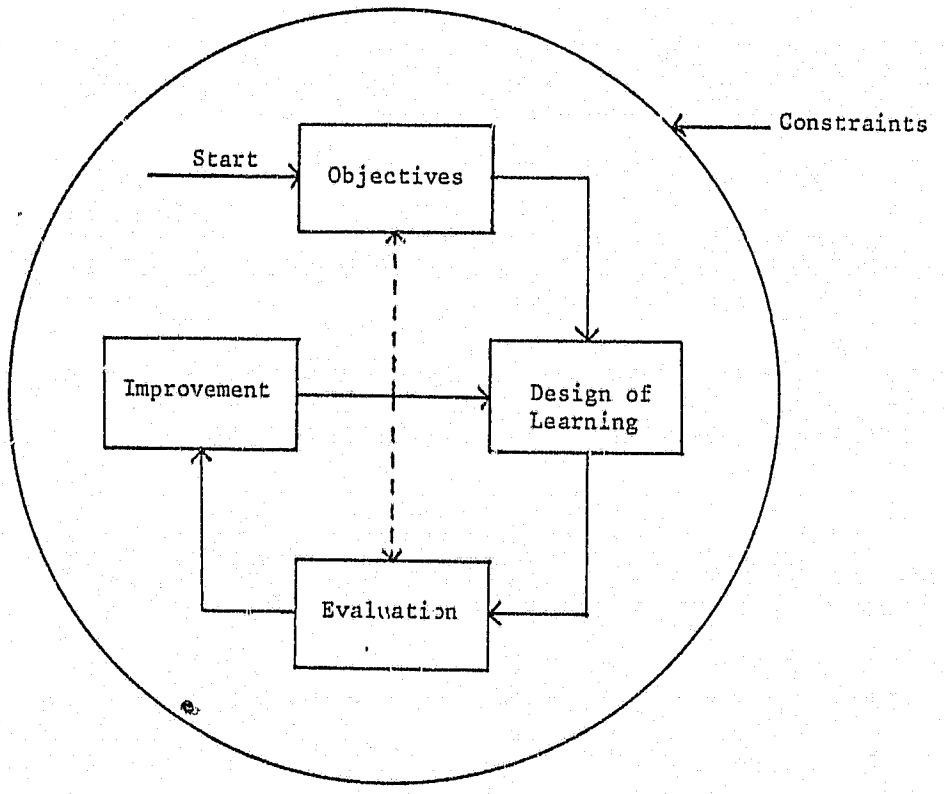


Figure 1: A Systems Approach to Education

Many see the systems approach as the development of the science of learning which is a very necessary development in today's world. Proponents of this approach claim that it has many advantages, some of which are increased learning effectiveness and efficiency and a system which is easily accountable for its success or failure. The entire process must be described in an orderly series of steps which are founded upon psychological theory and educational research. The instructional materials which are produced are empirically tested and revised when necessary. When the programme is put into operation its outcomes are predictable. The public therefore has an accountability system by which the degree of success achieved by educators in pursuit of the design objectives may be measured (Lomax, 1979(1)).

3.1.2 Some Criticisms

There has been some criticism and a rather large degree of apprehension as to how successful the systems approach can be in education. The main area of contention centres around the fact that many regard education as a human activity; an interaction between student and teacher which can often not be planned for in the systems approach. This interaction often involves exploration where the objective is unknown and a systematic, well defined approach precludes this sort of activity. Coles (1979) claims that 'education is a more or less unique interaction between teachers and learners, whatever its objectives.' He goes on to claim that in the systems model the means and the ends are prescribed by the teacher and the learner has to follow the teacher's route which prevents this unique interaction.

Macdonald-Ross (1975) states that the systematic approach is a way of learning which has been going on as long as recorded history with

regard to training. The rigidity of pre-set goals appears however to contrast with the teaching process when seen as an act of exploration by the teacher.

A number of the objections to systematic planning in general, and to the writing of objectives in particular, which are commonly cited can be easily dismissed, but the most substantial objection is the value system which they seem to imply (University Teaching Methods Unit, 1978). The taxonomy might be described as being predisposed to fit a pragmatic, materialistic, meritocratic view of society, in which few profound questions of value arise and in which explicit evaluation is always expected and preferred (Ormell, 1974).

Even though the systems approach has not met with universal acceptance as a means of attacking the education of the future, there is still much to be derived from adopting this approach and stating objectives. Piper (1976) states that he writes educational objectives to the extent that he believes learning ought not to be left to chance and because the organisation of learning experience is so complex that some form of systematised notation might improve the teacher's capacity to act rationally. This analysis could make the attainment of those objectives which we recognise and describe easier, and in so doing more time could be released for educational concerns of which we are less sure and less able to articulate.

One advantage of a systematised approach is that it facilitates the evaluation of teaching effectiveness. If it has been possible to define the objectives for a particular teaching sequence and criteria have been set which will indicate the relative attainment of these objectives, then the effectiveness of the teaching can be judged against the student's achievement or non-achievement of the criteria. One such

system will be described which was used in the construction and evaluation of the televised lectures.

3.2 A Criterion Referenced Approach

There is a broad spectrum of attitudes towards teaching. At the one end is the belief that the good teacher is quite literally born, not made, and there is very little that can be done to improve a teacher's proficiency. At the other end of the spectrum is the view that a good teacher is never born, but always made. All the views on this spectrum, with the exception of the extreme one first mentioned, must then agree that to some extent, a good teacher is made. Therefore there must be identifiable portions of the instructional act that are amenable to rigorous analysis and subsequent improvement. If this analysis is made by an individual teacher and decisions regarding a particular instructional sequence are made, the resulting quality of the teacher's instruction and the learner's achievements will be augmented if these decisions are made wisely.

A large body of research work exists which describes methods and procedures which can be adopted in order to analyse, evaluate and improve the teaching act (Baker, 1973). One of these instructional paradigms which has features which are common to many others will be described here. The approach to instruction is based on the central premise that the reason for a teacher's being in the classroom or lecture theatre is to bring about a change in the learners. Thus behavioural modifications can be considered the criterion by which instruction is judged.

3.2.1 The Ends

As the criterion of this instructional model is the behaviour of the learner, it is crucial to identify the intended behavioural changes at the outset of instruction. This will entail the stating of behavioural objectives which will become the ends of the instructional sequence. It is important to state these objectives in terms of observable student behaviour. In order to measure the quality of the instruction the attainment of these objectives will be measured. If there is no certainty as to what is being observed, no measurements can be made. Course aims are often confused with objectives. Aims are usually taken to be very broad statements of what a teacher intends to do and hopes his course will achieve; objectives are statements of what a student should be able to do (University Teaching Methods Unit, 1978).

Mager (1962) outlines very clearly how to write objectives in behavioural terms and lists several verbs, some of which are specific and some which are open to several interpretations:

non-specific	specific
to know	to write
to understand	to identify
to appreciate	to solve
to grasp	to compare

One of the difficulties encountered in writing specific objectives in terms of observable student behaviour is that the result is a list of test items. At least this is achieving the goal of defining what it is that the student must be able to do, but it will also mean that the task will begin to take on mammoth proportions. It is better to aim at writing objectives where the student behaviour is clearly specified but a generalization is made across a range of content. An example of this

would be: the students must be able to compute the correlation coefficient for any set of bivariate data for a maximum of 100 cases, the data being stated to a maximum of six significant figures.

Another important feature of a well-stated objective is that the criterion of acceptable performance must be stated. This is to facilitate a comparison between the actual attainment of the students and the required level. If all the students manage to attain the criterion then that instructional sequence can be considered to be satisfactory. This could be done in the form of a certain percentage of correctness or perhaps a certain time limit. When stating these minimally acceptable standards for the performance of the student, a student and/or class minimal level can also be stated, depending on the conditions of evaluation to be employed.

Finally, a well stated objective should also state under what conditions the behaviour is to be performed. This would include remarks concerning the students use of equipment (references, calculators etc) and any other facilities which are to be employed by the student.

Once these objectives have been defined they become the ends of the instruction. The question now arises as to which ends should be pursued. One of the greatest dangers of behaviourizing desired instructional outcomes is that those goals most amenable to measurement are usually the most trivial. There are several methods that can be adopted in order to avoid unimportant goals. One would be to consider the objectives in terms of the taxonomies of educational objectives.

Bloom (1956, 1964) has classified the cognitive and affective domains in terms of a hierarchy of behavioural objectives. A similar hierarchy exists for the psychomotor domain (Simpson, 1967). In the cognitive domain a hierarchy of intellectual behaviour is developed that ranges from knowledge at the bottom to evaluation at the top.

Similarly the affective domain ranges from receiving to characterization.

Once a set of behavioural objectives have been stated they can be classified according to this hierarchy. If the objectives always fall into the lowest category, it might suggest trivial objectives.

Bloom's work has been subject to frequent criticism, mainly on the grounds of lack of conceptual clarity (Ormell, 1974). The divisions between the levels in the hierarchies and indeed between the domains themselves are often regarded as artificial. Nevertheless they can be a useful means of deciding upon which objectives should be chosen. Other ways of classifying objectives have been suggested and Rowntree (1974) gives a summary of some of these.

A second method that can be used is the approach advocated by Tyler (1949). According to this approach a teacher would formulate a set of general objectives by consulting three sources: the student, the society and the subject matter. These sources are sometimes described as the social service view, the functional view and the cultural view (University Teaching Methods Unit, 1978). Once these general objectives have been formulated, they are then scrutinized in terms of two screens, the psychology of learning and the philosophy of education. The objectives that pass these screens are then behaviourized and used as the ends to which the instruction is geared.

A prerequisite before instruction is embarked upon once the behavioural objectives have been defined is that the pre-entry behaviour of the students is established in relation to these objectives. Thus some form of preassessment must take place. Any behavioural changes which are subsequently detected may then be attributed to the instruction, if these changes are detected during the instruction.

3.2.2 The Means

The next step in this criterion-referenced instructional paradigm is to decide on what instructional strategies are to be adopted. The actual strategy which is chosen is dependent on many variables and many learning principles have been enunciated as a result of experimentation in the psychology laboratory. It is not intended to discuss these here but merely to mention a few which have been found rather useful in the construction and development of the television lectures.

Popham (1973) puts forward the following four learning principles which he feels are of particular relevance in a criterion-referenced approach:

1. Revelation of Objectives. The students should be informed of the behaviours they will be expected to perform at the end of the instructional sequence.

2. Perceived Purpose. The student's perception of the purpose or value of the learning activity should be promoted. This would mean that the student is motivated towards learning.

3. Appropriate Practice. The student must have an opportunity to practice the kind of behaviour implied by the objective.

4. Knowledge of Results. The student should be given an indication of whether his responses are correct as soon as possible after he has made the response.

The concept of learning hierarchies has also been found to be useful, particularly in the analysis of the television lectures which is described in section 8.2. A hierarchy of component learning tasks is developed by beginning with a desired instructional objective and then asking in effect, 'To perform this behaviour, what prerequisite or component behaviours must a learner be able to perform?' For each behaviour so identified, the same question is asked, thus generating a

hierarchy of objectives based on observable prerequisities (Gagné, 1968). It was found that once the objectives had been behaviourally stated, this hierarchy was relatively easy to establish.

Finally, a taxonomy of learning conditions was used in some areas of the initial construction of the television lectures. A major effort to categorize tasks according to learning requirements or the conditions under which tasks are learned has been accomplished by Gagné (1977). He describes eight categories of learning: conditioned responses, the chaining of responses, verbal associations, sets of discriminations, concept formation, rule learning, the ability to use discriminations and rules and concepts to solve problems. These categories were however only used to a limited degree as their structure and implementation in the field of teaching statistics is still in the developmental stages (Chervany, 1977).

3.2.3 Evaluation Considerations

The final component of this criterion-referenced instructional paradigm is evaluation. Evaluation, as it is referred to here, is the assessment of the quality of the instruction. It is focused primarily on whether or not a pre-specified criterion has been achieved.

If the objectives are not achieved at the close of a particular instructional sequence, then there can be two principal explanations. First, the instruction might have been deficient. Secondly, the objectives might have been beyond the reach of the learners. Although both reasons should be carefully considered, in most cases it is the instruction which has been deficient. If this is the case, further analysis and revision of the instruction should take place; if the objectives are at fault, they should be changed accordingly.

This completes the general description of the instructional paradigm which was used in the construction of the television lectures. It could be represented diagrammatically as follows:

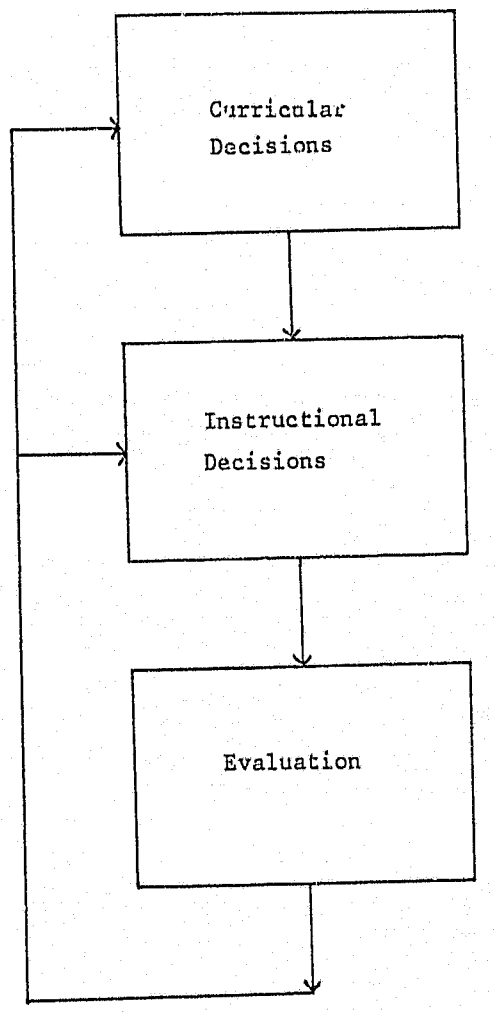


Figure 2: A Criterion-Referenced Instructional Paradigm.

The means of evaluation must also be given careful consideration. If a test is to be administered in order to assess the student's achievement, it must be a valid test. In other words the test must measure what it is supposed to be measuring if it is to give reliable feedback. Thus at this point the concepts norm-referenced test and criterion-referenced test will be introduced and discussed.

Glaser (1963) introduced the expressions norm-referenced measurement and criterion-referenced measurement when questioning traditional measurement practices. The essential difference between the two is that the one measures aptitude and the other achievement. A norm-referenced test is designed to ascertain a student's status in relation to the performance of a group of other students (the norm) who have completed that test. A criterion-referenced test attempts to measure what a person knows or what intellectual competencies an individual has acquired. Although it is convenient to think of these two types of testing purposes as totally distinctive, there is little doubt that they often overlap substantially in a given test.

It is precisely this overlapping, or rather the difficulty experienced in identifying, with any clarity, the two from each other that has caused the misuse of tests. A specific test should be used for a specific purpose. If the idea is to spread students out so that the best or worst can be identified, then a norm-referenced test should be used. If the idea is rather to determine how well students have mastered a particular behavioural objective, then a criterion-referenced test should be used. If a test is used for a purpose for which it was not designed, then difficulties in its interpretation will be encountered.

When instruction is to be evaluated, it is essential that a criterion test be used. If an instructional sequence has been designed to teach a well defined behavioural objective which is within the students' grasp,

then a criterion-referenced test will reveal how effective the instruction was in facilitating the attainment of that objective. A norm-referenced test will simply tell you where the students stand in relation to one another, and very little about the quality of the instruction. Although most tests have elements of both these concepts built into them, the following graph would illustrate their essential differences, even though these differences are somewhat exaggerated.

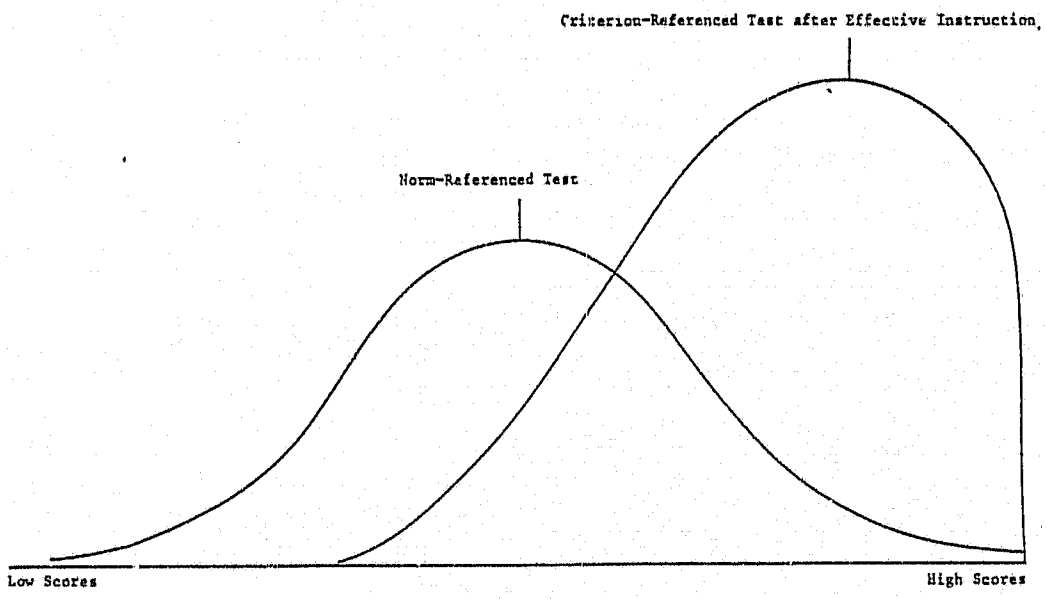


Figure 3: Criterion and Norm Referenced Tests

An instance where it has been claimed that inappropriate tests were used is that of the Coleman Report in the USA where it was revealed that money spent on schools did not seem to make any difference to their educational achievements (Carver, 1975). The result was that many legislators pushed for reduced expenditures in education as they reasoned

that more money spent on education did not appear to make any difference. Coleman, however, based his conclusions upon norm-referenced achievement tests which by their very nature, were insensitive to detecting the impact of effective instruction, even if it was present.

'...there are scores of instances in which hard-working faculty members have put together a really effective instructional program, only to be told that test results indicate the new program is not better than the old one. It is no better, of course, because its results were assessed by a norm-referenced achievement test- a test that tends to measure students' entry skills, not what they learn.' (Popham, 1978: 85).

How then is this pitfall avoided when assessing instruction? How is a valid and reliable criterion-referenced test constructed which is distinct from a norm-referenced test for a particular instructional sequence? The answers to most of these questions are still largely unknown. The technology of criterion-referenced test writing is still rather limited. The resolution of the assumed conflict in methodologies between criterion-referenced and more traditional tests has not yet emerged (Baker, 1973). More recently, a work has been published (Popham, 1978) which the author claims to be an attempt to write an introductory criterion-referenced measurement text, now that there is a sufficient body of relevant technology to share.

Difficult as the distinction may be between the two testing concepts, the criterion-referenced approach was borne in mind throughout the evaluation of the television instruction. This application is discussed more fully in chapter 7.

3.3 The Criterion Referenced Approach to Television Instruction

3.3.1 A Delimitation

The criterion-referenced instructional model which has been described is not the source or focus of this investigation. It was the most convenient model available which could be used as an aid in the evaluation of the teaching effectiveness of the television lectures. Its full application to the whole structure of the introductory statistics course is doubtful.

It is recognised that an evaluation model based on predetermined aspects of teaching and learning, especially those that specifically depend upon personal relationships and interaction (which is not usually the case when lecturing to large numbers of students), could ignore many of the most valuable and unpredictable aspects of teaching and learning. It is, however, also recognised that systematic analysis of the teaching act does not always mean that explicit evaluation will result, and often this analysis brings the teacher more quickly up against questions of profound value, than would have been the case if this analysis had not taken place.

A standard complaint about the systems approach is that, although it is not a bad idea, many proponents of the system do not usually recognise that a vast area must still be left open to the native judgement and intuition of the individual teacher (Macdonald-Ross, 1975). It is precisely for this reason that the criterion referenced approach was applied fairly rigorously only to the television lecture aspect of the course. The paradigm was very useful in other areas but they were far too complex to be subjected to the precise analysis which occurred in the television lectures. Besides attending the television instruction students can attend tutorial groups for the discussion of various

problems; they can (and do) consult lecturers privately; they are expected to complete a weekly assignment and they are also subjected to assessment type tests which could be classed as complete mixtures of the norm and criterion-referenced tests. The complete description and evaluation of the entire teaching system in this introductory statistics course would form an extremely intricate task. All interactions and learning tasks would have to be clearly defined and assessed. There are so many interacting variables that it is doubtful whether a study of this magnitude could be conducted with any meaningful results.

Thus the criterion-referenced approach was applied specifically to one teaching area of the course only: the televised instruction. The basic question which was then asked was: If televised lectures are taking place, are they responsible for bringing about any behavioural changes with respect to the learner? If they are not, then the whole exercise is simply wasted time and could be more profitably spent by both teacher and learner in pursuing other activities. However, if they are responsible for some behavioural changes, to what degree are these changes occurring? Can the televised instruction be optimized so as to be more effective in bringing about desired behavioural changes in the learner?

The evaluation of teaching in higher education has at least three areas of major concern: the assessment of lecturers; research into different teaching methods; and the evaluation of courses or educational systems (University Teaching Methods Units, 1978). The criterion-referenced instructional model is applied here to the first of these three areas: the evaluation of a lecturer's effectiveness in facilitating the achievement of desired behavioural objectives in the learner, via the medium of a televised lecture.

3.3.2 The Objectives of the Television Instruction

Although the entire introductory statistics course was not subjected to the instructional model described in this chapter, it did have certain influences in most areas of the course. One of these areas was the formulation of the general objectives or aims of the course. It is worth outlining some of the considerations which were taken into account in the formulation of the course aims.

The students that take this year-long course are all reading for a degree in commerce. The course is terminal in that the students are not required to proceed to a second course in statistics. The course content is orientated towards practical applications of statistical processes in business, commerce and industry. The entrance requirement is that the student must have passed the secondary school mathematics course at the matriculation level. The majority of these students can however be considered to be the more mathematically deprived from the group achieving a pass in mathematics at this level.

The course in general was designed around aims which were generated from these considerations. The aims of the television lectures were then generated in particular. A fuller enunciation of some of these aims will be given in Chapter 7. The objectives of each individual television lecture were then stated in terms of observable student behaviour.

It became evident that each lecture emcompassed the mastery of several behaviours. Each individual lecture was therefore broken down into segments or separate instructional sequences, where each sequence was linked to a specific objective. A list of these objectives for each of the six lectures can be found in Appendix I.

A minimal level of performance was then decided upon by mutual agreement of a team of lecturers responsible for lecturing the course. A desirable standard was considered to be at the 80% level. Thus for

the lecturing to be considered effective in any instructional sequence, 80% of the students must have achieved the desired behavioural objective. However, once all the circumstances were considered, an absolute minimum level of 70% would be the cut off point for an instructional sequence to be considered in any way 'effective'. This level (70%) was decided upon in the first instance, with the view that it should be increased once any deficiencies in the instruction were detected and rectified.

The subject matter covered in these six lectures was an area that had not been encountered before by the students. Thus their pre-entry behaviour was not established with respect to the objectives of the lectures. A pre-test to determine the students' mathematical ability was administered, but this was not essential to the experimental design as is indicated when discussing the statistical analysis in section 8.3.

The instructional strategy to be employed was decided upon by the lecturer, according to what he considered to be the most suitable to his style, personality and capabilities. The learning principles as outlined in this chapter were also considered when the instructional strategies were chosen. These will be amplified in Chapter 6.

The evaluation of the television instruction forms a main area of the results of this experiment and is discussed in section 8.2.

4. THE STRUCTURE OF THE TELEVISION PROGRAMMES

When the 'live', traditional lecture was to be transposed onto videotape several decisions had to be made with regard to how this transposition was to take place.

In the first instance, was television the best medium to use in order to replace the live lecturer? Would not a tape-slide sequence be just as, if not more, effective in this subject. Another decision which had to be made was one regarding the form of the lecture. Should the lecturer be recorded in his usual environment, holding forth from the lecturing platform in between scribbling notes on the chalkboard? Or should this 'natural' format be changed?

These and other questions are discussed in this chapter while outlining the structure which was developed and used in the television lectures produced for this study. Many of these decisions have strong cognitive overtones. These areas are identified and relevant research is acknowledged. These areas are not however being researched in this study. The feature of the television lectures which was essential to the research is then outlined and its intended use is specified.

4.1 The Development of the Television Programmes

4.1.1 Initial Philosophy

Some research has been done related to media selection which specifies the relevant variables in terms of the attributes of the media rather than in terms of the media themselves. Attributes of a medium

are the capabilities of that medium to show objects in motion, objects in colour, objects in three dimensions; to provide printed words, spoken words, simultaneous visual and auditory stimuli; to allow for overt learner responses or random access to information (Levie and Dickie, 1973). The rate of development of educational technology would make it difficult to compile a comprehensive and detailed taxonomy of media attributes which are critical in media selection but Tosti and Ball (1969) have made an admirable attempt and offer six 'dimensions of presentation'. If it is decided (once media attributes have been considered) that more than one medium is capable of providing the required attributes, the choice could then be made on the basis of pragmatic determinants such as cost.

These media attributes were considered when the original choices with regard to the medium and format of the lectures were made, although it must be stated that pragmatic considerations such as available resources and cost played major roles in these decisions.

It was decided to televise the lectures and the format would reflect a slight change from that of a direct recording of a traditional 'talk and chalk' lecture. The lecturer would be seated at a desk in his office for the major part of the presentation. Occasional use could be made of a chalk board on the wall of the office if so desired. The major portion of the information which is usually placed on the chalkboard/overhead projector during a lecture would be prepared in graphic form beforehand and revealed on the monitors as and when required.

These initial decisions were not seen as attempts to generate something 'extra' from the new medium, but simply to ensure that the lecture was on an equal footing with that of the live presentation, at least as far as the supply of information was concerned. Ksobiech (1976), while admitting that there are some opponents to the view, claims that

there is no conclusive evidence to indicate that dynamic production techniques (such as animation, colour, special effects, etcetera) increase the amount of information gained by students. This was the view adopted when secondary decisions were made after the medium and format had been finalised.

The programmes were produced in colour as this facility was available and the live lectures are performed in colour. This decision was also based on research which has compared colour with black and white pictures and has usually found no significant differences in learning; when results did favour one or the other, this occurred with about equal frequency (Levie and Dickie, 1973).

Some animation and special effects were occasionally used in conjunction with the graphics, but these were equated to the 'live' lecturer developing or highlighting points on the chalkboard. The transition was therefore an attempt at replacing the 'live' lecturer with an equivalent form of a lecture. This attempt, in terms of these limited aims, could be said to have been 'successful' as has been outlined in section 2.3.

There was however one important distinction made between the 'live' and 'televised' lecturer. The television personnel insisted that their most successful programmes had an optimum length of approximately fifteen minutes and therefore the introductory statistics programmes should have a maximum length of twenty minutes. This point was taken and the strategy was that the theoretical content of the course would be presented by the subject 'expert' for part of a forty five minute lecture period. The remainer of the time would involve practical applications of the theory and further clarification on some areas of the television lecture. Other lecturers or senior tutors would be used for this second task.

However, when the first of these lectures was replayed, it soon became apparent that the attention period of the students was not fifteen minutes. Attention, as it is referred to here, is generally conceived of as an activity which must occur before a stimulus will be associated with a response (Glaser and Cooley, 1973). This 'attention', although present at the beginning of each televised lecture, lasted for a maximum of about five minutes (varying from student to student) before a 'failure of attention' occurred. This was determined by lecturers observations from the front of the lecture theatre during the replays and was loosely interpreted as an obvious action by the student, such as gazing for a period of time at areas in the room, other than the monitors, without any response to the information being transmitted by the monitors. This was further evidenced by student's complaints of 'getting lost'. Once attention was re-established the student often found that he was confronted by a unique mathematical expression which was meaningless. Once this occurred the student invariably failed to derive any further benefit from the lecture and the frequency with which 'failures of attention' occurred increased.

Thus the following dilemma arose: either the televised lecture had to be decreased from twenty to about five minutes in length or some means of holding the students 'attention' for a longer period of time had to be found. The first alternate was rejected as totally impractical and the second was pursued.

4.1.2 The Optimum Length

One of the major complaints received from the students was that they experienced difficulty when viewing mathematical expressions on the monitor. In some instances the expression warranted their total attention while some explanation with regard to it was being propounded,

while in other cases some note of the expression was needed for le reference, before it disappeared from the video monitor. Thus indecision often resulted as to whether to take notes or to direct full attention at an explanation taking place. This appears to agree with research which claims that learning discrimination involves two processes: first, an attentional response to the relevant features of the stimulus display and secondly, the correct response to the relevant feature (Zeaman and House, 1963). It is claimed that an important aspect in instructional design is the development of materials that facilitate attentional learning on the part of the learner (Glaser and Cooley, 1973).

In an attempt to overcome this difficulty experienced by the students it was decided to change the background colour of the graphics. One colour was used when the students were expected to direct their full attention at the monitor. This background colour was changed when they were expected to make some note of the graphics appearing on the monitor.

This procedure had an unexpected result. Not only did it appear to solve the 'attention' problem but it meant that at certain points throughout the lecture (which could be controlled) the students were responding actively to the lecture. These 'response points' acted as 'attention getters' and the frequency with which complete 'failures of attention' occurred, dropped dramatically. It became possible to produce a televised lecture of forty five minutes duration, using this technique, without any complaints from the students or noticeable 'failures of attention'.

Thus no optimum lecture length was stated. It could be varied to suit the circumstances without an observable change in effect.

4.1.3 Coloured Backgrounds

Background colour refers to the colour transmitted by the monitor when graphics are used in the television lectures. Any lettering or mathematical expressions on the graphic are superimposed on this colour. Research has been conducted on background colours in an attempt to establish the effect these colours have on visual acuity (Snowberg, 1973). Snowberg's study revealed no simple conclusions as visual acuity appears to be related to the luminance and transmission of the colour concerned. Some small differences were detected with white being recommended as a background colour (black lettering was used in this experiment) and blue backgrounds were not recommended when critical legibility was required.

Snowberg's study was conducted using colour slides and as the findings were not sufficiently conclusive the recommendations were not adhered to for the selection of background colours for the television lectures. A far more pragmatic standpoint was adopted. When reproducing copies of video-tapes, there is a resultant loss of quality with each generation. Certain colours deteriorate in quality more rapidly than others. For example, red becomes very 'scratchy' as one goes from a second to a third generation tape with the equipment which was used in this experiment. Therefore background colours were chosen for the television lectures which gave a lower 'deterioration of quality' rate when copies were made from the master tape.

When the student's full attention was required, a blue background was used. If he should have made a note of whatever appeared on the monitor, a green background was used. White lettering appeared with both the above background colours.

Brown was used as a third background colour to elicit another type of response from the students. This development was introduced when producing the six television lectures for this study and is discussed in section 4.3.

4.2 Some Cognitive Considerations

Ksobiech (1976) has stated that one of the major conclusions which research in the area of instructional television has yielded is that there is no significant difference in learning between instructional television and conventional classroom instruction. Moldstad (1974) has indicated that this conclusion has positive value for school administrators who can thus justify the use of instructional television to solve problems in the areas of personnel, scheduling and, to some degree, finance.

This is the bland attitude mainly adopted in this study. The subject of the investigation is not the comparison of television instruction with some other medium. It is rather a measurement of the instructor's effectiveness which is manifested in student behaviour and an investigation into the optimization of this effectiveness. This is all accomplished through the medium of television. The type of medium is not in question; that has already been accepted.

This does not mean, however, that the cognitive considerations of how learning from television occurs are considered unimportant. On the contrary, this area requires far more research in the future, yet up to now very little attention has been paid to the question of how, if at all, the symbolic modes of visual media affect cognitions (Cohen, 1979(2)). Many of these considerations are present in the television lectures in this study and some of these will be mentioned briefly.

One question which is of concern is the amount of detail which is present on the television graphic and how this affects student learning. Research conducted in this area is summarized by Dwyer (1970). One general outcome which is of particular interest is that in presentations where the pace is fixed, as is the case with these television lectures, visuals containing little detail tended to be more effective. Dwyer speculates that in fixed-pace presentations learners do not have time to take advantage of the added information available in realistic visuals and may be distracted from important verbal information. When learners can take as much time as they wish, realistic detail may provide additional learning experiences.

The cognitive influences regarding the use of colour also need further consideration. Although Kanner (1968) reviewed eighteen studies dealing with the effectiveness of colour in television and concluded that the addition of colour does not result in greater learning (even though colour is almost always preferred by learners), colour may enhance learning when it is used to emphasize relevant cues and aid in making appropriate discriminations (Norman and Rieber, 1968). In a summary of the state of the art Kanner (1968) states that there is a scarcity of information on the topic of colour and human learning. This view is supported in a study of colour's effectiveness in facilitating learning in the affective area (Booth and Millar, 1974), while Dwyer (1976) has attempted to relate I.Q. levels with the effectiveness of black-and-white and colour illustration.

The motion aspect can be, and sometimes was, introduced into the graphics in the television programmes is another cognitive consideration. It appears that the functions which motion can perform to enhance learning have not been fully identified. Television combines many of the attributes available in other media and theoreticians concerned

with this type of communication are particularly aware that when media attributes are used in combination they do not operate independently; the use of an additional attribute such as motion may alter the functions which other attributes may perform (Levie and Dickie, 1973).

More recently, research into the effect which television formats of representation have on mental skills has taken place (Cohen, 1979(2)). The close-up, split-screen, slow motion and time-compression are examples of what is meant by formats of representation. Then, of course, there is the consideration of how learning takes place in mathematically based subjects. There are several controversial theories which attempt to describe how mathematical learning actually takes place (Shulman, 1970).

This, however, is not a psychological study. Thus the cognitive considerations which have been mentioned and which are present in the television lectures, although considered as important and always borne in mind, will not be investigated any further.

4.3 The Means of Evaluation

4.3.1 Intermittent Questions

This study is concerned with the cognitive level which students achieved during the television instruction. In order to determine this level of achievement, questions were inserted into the television instruction. A third background colour was introduced to produce another type of response from the students. Each of the six television lectures which were produced for this study was broken down into distinct instructional sequences, where a specific behavioural objective is associated with each sequence, as outlined in sub-section 3.2.2. In order to determine whether or not a student had achieved a specific objective, a question which was related to that objective was inserted at the end

of each instructional sequence. A brown background colour with white lettering was used for these questions.

The technique of inserting questions at intermittent stages during television instruction has been used before (Gropper and Lumsdaine, 1961) when developing programmed instruction television lessons. The effectiveness of the programmed television lesson was then determined by student performance on achievement tests. The questions inserted in the television lectures in this study were formulated on a criterion-referenced test approach and were designed to determine whether or not students had satisfactorily achieved the behavioural objective that had been specified for a particular instructional sequence.

An added contribution which these questions made to the television lectures was that they acted as additional 'attention getters'. The students were expected to actively respond to the questions by writing down the answers. Thus when green and brown backgrounds appeared the students responded actively and therefore participated during the television lecture. All six programmes were relatively long (between forty and fifty minutes) yet 'failures of attention' were hardly ever observed. This appears to correlate with studies which have investigated the role of questions in maintaining attention to textual material.

Rothkopf (1970) coined the term 'mathemagenic behaviour' to describe attending behaviours which give birth to learning. In studies of attention and learning from continuous material it was found that test-like questions presented before or after the material to be learned, are mathemagenic. Both prequestions and postquestions facilitate learning but to different degrees. Prequestions tended to narrow the range of attention by providing the individual with a criterion for acceptable behaviour. Postquestions on the other hand, faciliated both question specific and general learning, because attention is paid to the

whole passage, not a particular stimulus within the passage (Bull, 1973).

The insertion of these questions was not primarily intended to increase the effectiveness of the television lectures but rather as a means of evaluating the effectiveness of individual instructional sequences, although both these aspects appear to have been achieved. In addition to this a majority of students indicated a preference for the lectures with inserted questions.

4.3.2 The Value of Responding

Once these questions were asked, the students were expected to respond and record their answers for later reference. The questions were revealed for the second time at the conclusion of the lecture for a few moments, with the correct answer displayed underneath each question. Thus there was a relatively small delay in feedback once the students had made their responses.

As has been stated, the primary motivation for inserting this question-response procedure was not to improve the effectiveness of the lecture. An investigation was not conducted to compare this type of television lecture with another type, although there appears to be substantial evidence to suggest that the procedure does improve the instruction.

Levie and Duckie (1973) state that it seems reasonable that learning should be facilitated by giving the learner practice in what he is expected to do and informing him of his success or failure. A key issue that has concerned researchers interested in response and feedback variables involves the relative effectiveness of overt responding versus covert responding. Overt responding usually refers to writing answers or selecting from multiple choice alternatives while covert responding involves 'thinking' the answer or just reading. Levie and Dickie claim that overt responding facilitates learning only when the prescribed re-

sponses are relevant to the criterion objectives and is particularly effective when the required response is not already in the learner's repertory. It tends to facilitate learning more for difficult material than for easy material while covert responding is better for fast presentation rates.

It appears then that when the students knew that postquestions were to appear at the end of each instructional sequence and they were expected to respond overtly, the 'attention level' was heightened and learning faciliated to a greater degree.

The sample of students who participated in the evaluation of the television instruction were involved in a slightly different response-feedback mechanism which is outlined in section 5.3.3.

4.3.3 The Primary Intention

It can be argued that if the lecturer is going to spend his and the students' time in the lecture situation then some benefit in terms of learning should be gained by the students. Thus if a lecturer is going to spend five or ten minutes during a lecture in an attempt to impart some body of knowledge, or concept, or whatever, then it would be as well to examine whether the students have achieved what it was they were supposed to have achieved during that time. This is the primary objective in inserting the questions at the end of each instructional sequence of the television lecture.

The questions themselves were very basic to the instruction and were kept as simple as possible. If the student has grasped the basic essence of what it is that the lecturer has attempted to convey, then the future possibilities are almost limitless. However, if a student has failed at this early point, as will be indicated by these simple, basic questions, then the probability is rather high that he will progress no

further. If the overall result indicates that the majority of students failed, at this point, to achieve th. objective then there is a strong possibility that the fault lies with the instruction and this information should be used in an attempt to improve the instruction. An important principle of learning is that an instructional environment should be designed so that information about the student is provided to the student and the teacher so that each can appropriately use this information for the design c. subsequent instructional activities (Glaser and Cooley, 1973).

Cohen and Brauer (1969) suggest that student achievement of learning objectives is the main criterion on which studies of faculty and of instructional efforts should be based. In their view the use of student gain on short-range objectives as a measure of teacher effectiveness is generally acknowledged as being more valid than the use of such criteria as, for example, the teacher's e t expended or the various perceptions of observers. This is the attitude adopted in this study and the students' gain on short-range objectives was measured via the intermittent questions in the lectures. No attempt to measure the achievement of these objectives on a longer term basis was attempted. If a day, week or month is allowed to elapse before some further assessment is made, an almost infinite number of variables and interactions are introduced. This assessment is then some type of measure of the quality of the entire course and is no longer a measurement of the effectiveness of the lecturer's instruction. The results then become very difficult to interpret.

It is therefore important that once the questions have been asked, the responses are immediately recorded, before any other variables or interactions become involved. A student who is unable to answer a question immediately due to the poor instruction preceeding the question

might be able to answer the question before the lecture is over because of some other instructional sequence which was particularly effective and clarified the existing difficulty or because of some assistance rendered by a fellow student in the course of the lecture. This attitude is reflected by Popham (1971) when he states that if the student's content knowledge and ability level can be held constant, then a teacher's general teaching ability can be measured by testing his students for the skills specified by the teacher's instructional objectives.

This approach poses certain serious methodological problems like finding topics of instruction unfamiliar to the students that can be taught in a relatively short time but not be confined to only low level behaviours. This difficulty was considered and was instrumental in the topic choice for the six lectures in this study which not only provided content with which the students were unfamiliar but also with fairly high behavioural levels.

During a particular instructional sequence lasting for only a few minutes, a students knowledge and ability level is almost constant (although this is not usually the case for an entire lecture). Therefore any behavioural changes which occur during this sequence can be attributed to the effectiveness of the instruction. The immediate collection of student responses at the end of each sequence is therefore of paramount importance and is the central feature in the evaluation of the lectures in this study. This was achieved with computer assistance and the method is fully described in section 5.3.

These questions therefore provide the lecturer with a step by step analysis of his effectiveness. As only component parts are classed as good or bad, and not the entire lecture, the lecturer might be more amenable to subject himself to this type of evaluation. Few lecturers would claim to be fully effective in every sequence of every lecture

they deliver. The evaluation of lecturer effectiveness is of great importance in tertiary education. However, most lecturers are very sensitive to this sort of thing, partly due to the uncertainty of what it is that is being measured.

'Not unnaturally teachers are highly sensitive to any form of assessment of their work. This sensitivity can be eased if these points are recognised. Nevertheless, whether they like it or not, teachers are assessed. They do seek promotion and they do apply for new jobs. The question is not whether teachers in post-secondary education should be assessed but how and when, and what for? If teaching is a major function of post-secondary teachers it seems reasonable to suppose that the effectiveness of their courses should be a major consideration of promotion boards.' (Bligh, 1975, p213)

5. A COMPUTER BASED TEACHING SYSTEM

Glaser and Cooley (1973) state that the character of various epics in history has been determined in large part by the tools available in the period and in an analogous way the character of teaching is determined by available tools; the number of available tools for teaching is increasing daily in the age of 'multi-media technology'. They go on to claim that an important task in this area is to design instructional instrumentation that is responsive to the learner's performance so that information is supplied with minimum delay. An important feature in the experimental design for this study is the ability to record student responses as they take place at the end of each instructional sequence. Due to a recent computer system design development this feature is easily facilitated. The system was designed and developed in South Africa over the past three years and represents yet another teaching tool which is available to us.

Three major areas of computer innovation have been identified by Hatfield (1969):

(a) computer-assisted instruction (CAI); this includes drill and practice systems, tutorial systems and dialogue systems,

(b) computer-managed instruction (CMI); the computer is used here to assist the teacher in effectively administering and guiding instruction,

(c) computer-assisted problem solving (CAPS); in this category the students write computer programmes to solve selected problems and in this way gain insights into certain concepts.

The computer system used in this study cannot easily be slotted into one of these three categories as it attempts to cover almost all three. The system's full name is the Cybercom Multimode Education Communication System and is usually referred to as the Cybercom System (Lomax and Bayne, 1978). As the name indicates it attempts to operate in a variety of modes so as to accommodate several teaching situations. A description of the structuring of the main components of the system is given, followed by an outline of the various modes in which it can operate. Finally the mode of operation used for the evaluation of the television lectures will be expanded upon with details of how the relevant data was collected.

5.1 The Structure of Components

The Cybercom Computer System has a basic structure where individual computer access is available for each student and the teacher in a classroom/lecture theatre situation. This is achieved by providing the students with a simplified and smaller version of the 'normal' type of computer terminal which usually consists of a typewriter keyboard and video-display unit, which is referred to here as a student panel. The teacher/lecturer/instructor is provided with a more sophisticated computer terminal which consists of a typewriter keyboard, a video-display unit (VDU) and a panel of mode control buttons, which is referred to as an instructor console (the term instructor will be used to designate a teacher, lecturer or tutor). It is via these 'terminals' that the instructor and students achieve communication with the computer and each other. A mini-computer is used to facilitate this communication and the relationship between these three basic components is indicated in figure 4 as A,B and C. Two more units, a printer/reader unit, D, and a disk unit, E, make up the remainder of the

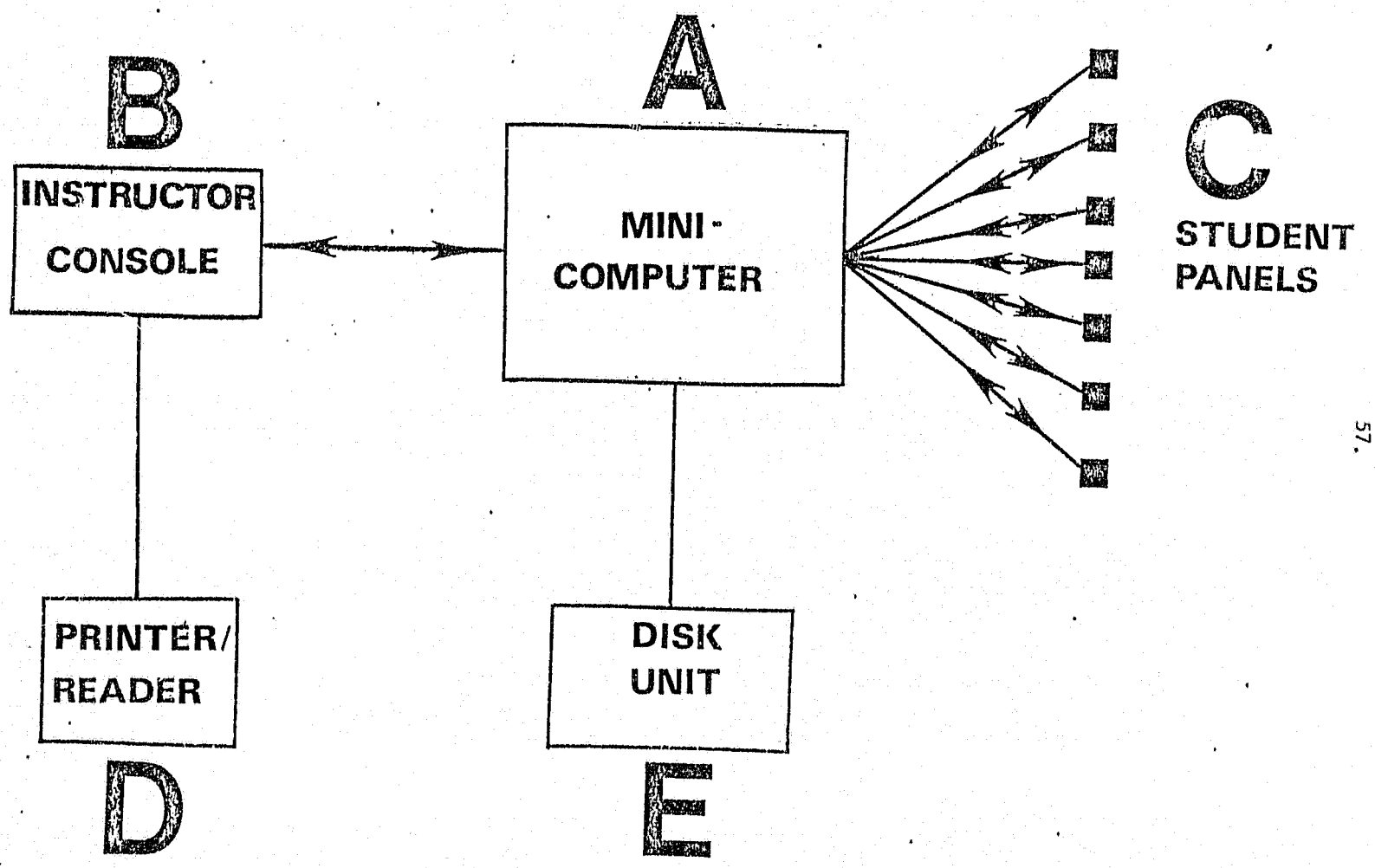


Figure 4: The Structure of Components in the Cybercom System.

hardware components of the system.

A description of these units will be given in this section which outlines their basic functions (Zawels, 1976).

5.1.1 The Mini-Computer and Disc Unit

The mini-computer, disc unit and some central electronics which control the communication between the various components are housed in a cabinet which is stored remotely, out of the confines of the lecture theatre. It is useful to have these units within easy access of the teaching area as the instructor may wish to use the disc unit (which reads from or writes to diskettes). The disc unit is an optional feature and can be used to provide additional storage for the system or rapid loading of large programmes.

The system has a capacity which allows a maximum of 256 student panels to be connected at any one time. These panels can operate in any of eight differently assigned groups. The lecture theatre which was used for the evaluation of the television lectures had a Cybercom system installed where the projection booth at the rear of the theatre was used to house the mini-computer and disk unit.

5.1.2 The Instructor Console and Printer/Reader

The instructor console consists of three components: (a) The typewriter keyboard, (b) the video-display unit and (c) the panel of mode control buttons. The entire console is indicated in figure 5 and the three component parts are more clearly identified in figure 6. The three components will be discussed individually.

(a) The typewriter keyboard: this is similar to the keyboard of an electric typewriter , although some special function keys have been added. The instructor may enter lesson programmes and/or control the

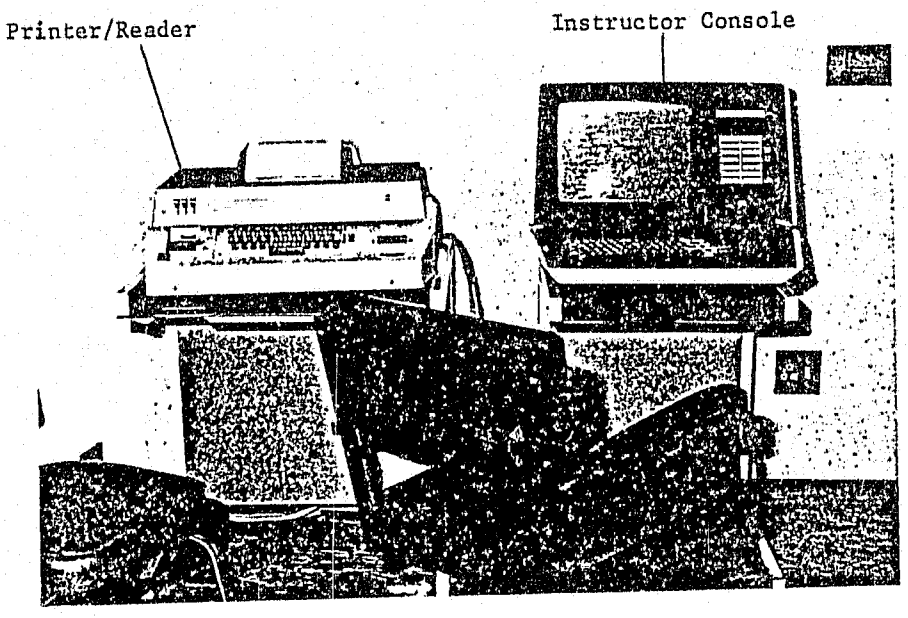


Figure 5: The Instructor Console and Printer/Reader

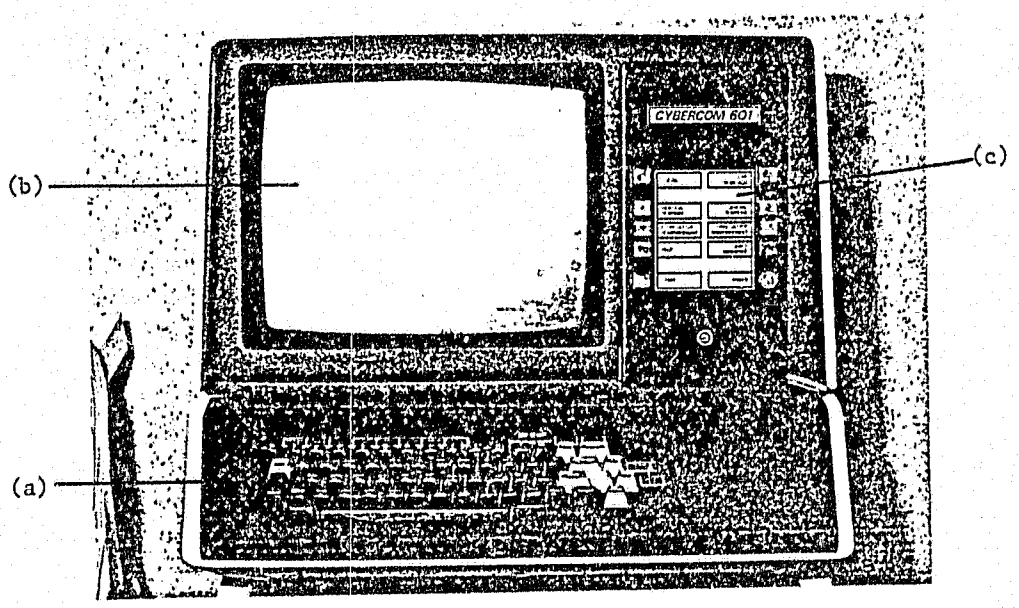


Figure 6: The Three Components of the Instructor Console

general proceedings as he so desires before, during or after any form of teaching which has taken place, by using this keyboard.

(b) The VDU: This component displays any information which the instructor and/or the students have entered into the system in addition to any analysis of this information which the instructor may ask for.

(c) The panel of mode control buttons: These buttons are used to insert the system into a particular mode of operation (illustrated in figure 7). The various modes of operation are discussed in section 5.2.

An additional feature which is available for the instructor is a printer/reader unit which is shown in figure 5 alongside the instructor console. The printer will provide a hard copy of any information which appears on the VDU. A paper-tape reader and writer is attached to the printer and enables the instructor to produce and feed into the system copies of lesson programmes. This obviates the need to type the lesson programmes into the system via the keyboard.

These two units were positioned at the front of the lecture theatre which enabled the instructor to make use of them at any stage during the lesson.

5.1.3 The Student Panels

The student panels are no larger than an average sized book (24cm×21cm×3cm). Figure 8 illustrates their relative size when they are used in the lecture theatre situation. The panel consists of a keyboard in a flat plane and a small viewing window. These features are illustrated in figure 9. The keyboard is similar to that of a typewriter with a few special function keys. The student communicates with the system by gently pushing the relevant key until a 'bleep' sound is heard. The symbol which the key represents then appears in the panel window. This window will view the four most recent symbols which have been entered into

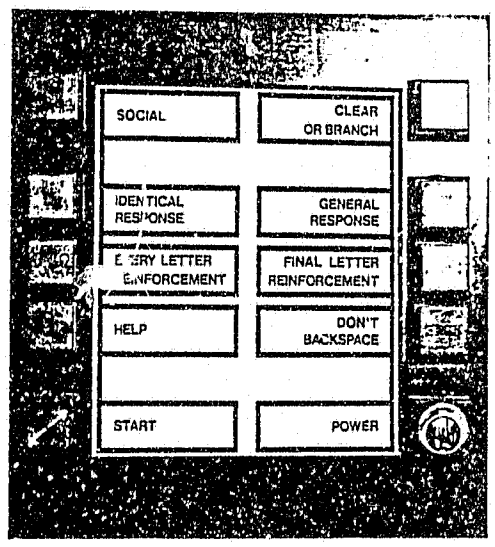


Figure 7: The Mode Control Buttons

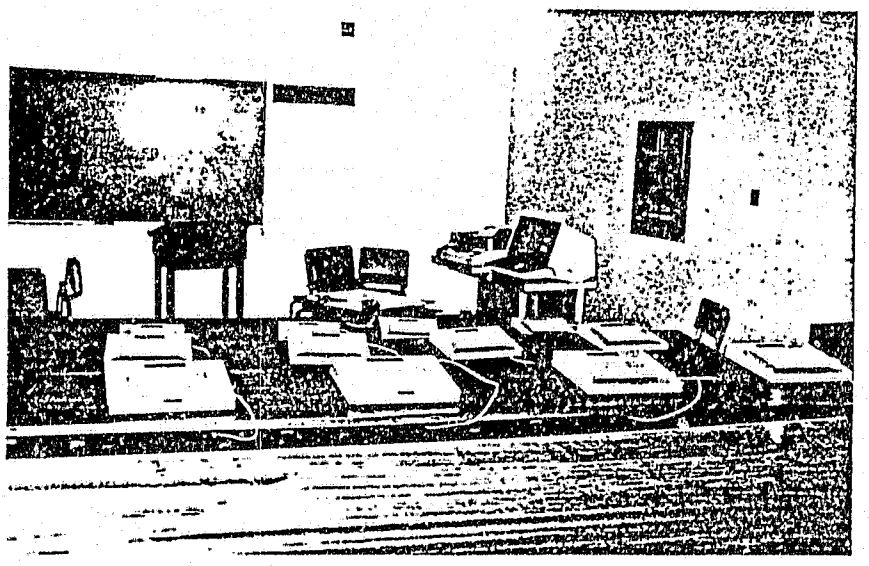


Figure 8: Student Panels

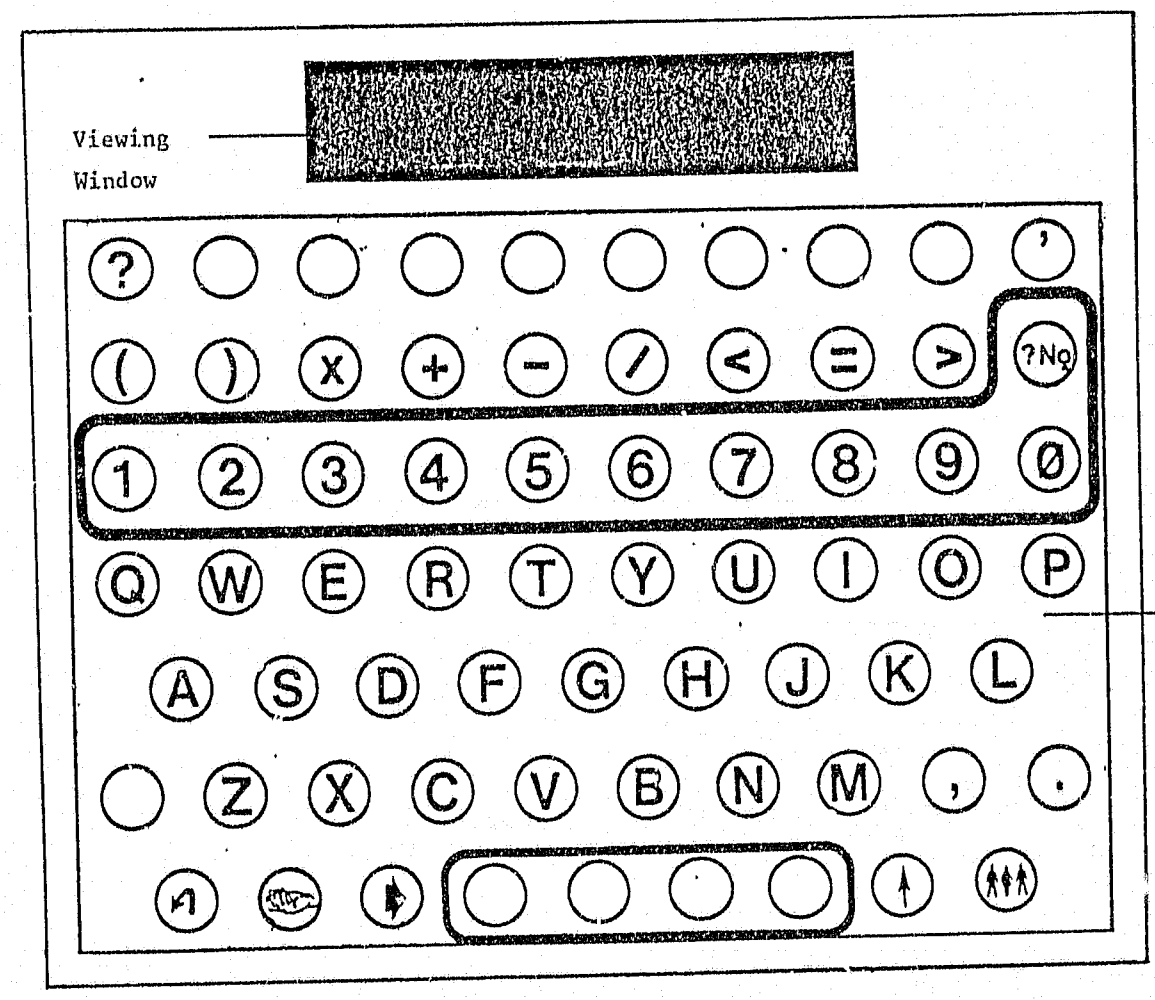


Figure 9: The Student Panel

the system. The following example will illustrate this feature.

If the number one is entered the window will reveal:

1

If the numbers two and three are then entered, the window will reveal:

1 2 3

If the numbers four and five are then entered, the window will reveal:

2 3 4 5

Thus the student views only a certain part of the information which he has entered. He may move along in either direction in order to view information.

The student receives information by means of messages which appear in the viewing window or by means of two lights situated in the viewing window which flash at certain speeds when supplying certain information. For example, if a student answers a question by entering the answer into the system, these lights will flash with a very high frequency if the answer is correct. Their position in the window is indicated below:

X X

The interaction of these components enables the instructor and the students to be in constant communication with each other, whatever teaching situation is taking place. Not only can the students monitor the instructor's behaviour, which is usually the case in a traditional lecture, but the instructor can monitor every student's behaviour at every point in the lecture.

5.2 The Modes of Operation

The Cybercom system can be inserted into different modes of operation by the instructor, by using the mode control buttons on his console. The two basic modes of operation will be outlined in this section in order to illustrate the uses to which this system could be put in a teaching situation.

5.2.1 The Social Mode

If the system is employed in the social mode, then all the students react simultaneously to some stimulus. The two main teaching areas where this mode could be used are the lecture and discussion.

In the lecture situation, the instructor is usually involved in imparting knowledge and understanding to the students. This is probably the most basic and commonly used teaching situation. Questions are often asked at certain stages of the lecture in order to determine whether or not students have understood a certain explanation. The system now allows the instructor to view the answers of all the students rather than one or two, because as the students enter the answer into their panels, these answers are displayed on the VDU. This was the situation employed for the evaluation of the television lectures. The instructor can then decide whether he or the system will inform the students of the correctness of their answers (if reinforcement is desired).

The discussion is another area where the social mode can be used. The instructor or discussion leader can monitor all the responses to certain points raised and can direct or lead the discussion in a manner which is based on this information.

5.2.2 The Individual Mode

If the system is inserted into the individual mode, then students can proceed through a situation at their own pace. Two areas where this mode would be employed is in self-paced learning and testing or grading.

If a student is progressing through a programmed instruction text, the answers can be previously inserted into the system. The student can then interact with the system in order to obtain reinforcement at certain points in the learning sequence. Students can be placed into different groups, depending on their rate of progress. A special help feature can be activated by students who are experiencing difficulty. The system keeps a record of every student's progress, which is available for the instructor. In this sense, the system is being employed in a CMI manner in addition to CAI.

Secondly, the system can be employed for the recording and grading of test answers. Students would insert their answers into the panels which would then be recorded and marked for the instructor. The flow-chart in figure 10 indicates the mode choices which the instructor has when using the system.

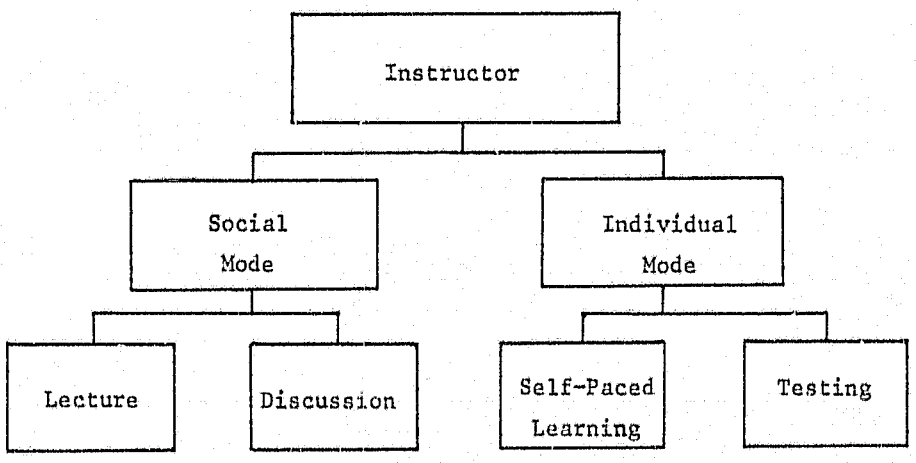


Figure 10: The Mode Choices

5.2.3 Results and Analyses

In addition to storing all information which is entered into it, the system will produce a battery of reports and analyses almost immediately on request by the instructor. This provides the instructor with details of the group's performance in addition to information about an individual's progress. These reports and analyses can be requested at any time during a lesson. A hard copy of the final report and analysis can be produced by the printer at the end of the lesson for later reference. Examples of some of these reports will be given in the next section when the use of the system in analysing the television lectures is discussed.

5.3 The Evaluation of the Television Programmes

The Cybercom system has now provided a means of collecting students' responses as they occur in a normal lecture situation. This provided a strategy with which to evaluate the television lectures. The television lectures were shown in a familar setting to the students. However, when a question was asked following a particular instructional sequence, the students would respond by entering the answer into the student panel in addition to making a note of it for later reference. This gave the lecturer immediate feedback as to how effective that particular teaching sequence had been in bringing about some predetermined behaviour.

In order to achieve this, a particular operating mode of the system was chosen, and a special type of use was made of that mode which was necessitated by the experimental design. This operating sequence is described in this section with accompanying illustrations of the type of feedback which was received at various points during the lecture.

5.3.1 The Lesson Programme

Before the commencement of each lecture, the answers to the questions to be asked during the lecture had to be fed into the system. This is easily facilitated via diskette or paper tape. However, the content of these answers deserves some attention. The information which is fed into the system is known as the lesson programme. The reason for this programme being entered into the system prior to the lecture is that the system must have a basis from which it can compare the responses received from the students and then indicate what percentage of these qualify as correct responses. This will be more clearly indicated if we consider a particular lesson programme.

The first of the six lectures used in this study, lecture M35, has a lesson programme as indicated in figure 11.

```
                    @EHI
00 M35
@I
01 2\TWO
@I
02 6\SIX
@I
03 52\FIFTY TWO
@I
04 X=4\4\FOUR\X=FOUR
@I
05 0\ZERO\X=0\X=ZERO\NOUGHT
@I
06 3/8\0.375\0.375
@I
07 1/16\0.0625\0.0625
@I
08 1/8\0.125\0.125\F(3)=1/8\F(3)=0.125\F(3)=0.125
@I
09 DISCRETE\DIS
@I
10 NO
```

Figure 11: The Lesson Programme for Lecture 1-M35

The first line, where the symbols @EHI appear is simply an initialization process performed by the system at the commencement of each lesson programme. The number 00 is the code for the name of the programme: M35 appears in this position.

The third line has the symbols @I. This signifies a condition which has been chosen for the form of the answer and means that the answer which the students enter must be identical to the answer which is given for question one in line four. If this condition was not stated then a general answer which contained a few key words or phrases would be accepted as correct. This facility exists for answers which are difficult to quantify. However in this subject most answers are quantitative and therefore the 'identical' condition was chosen.

On line four the number 01 appears which represents question one. The answer to this question is two. As students could indicate this answer in more than one form, alternatives are given. In this case they are indicated as 2 or two and programmed as 2\TWO.

This pattern continues for questions two to ten. In all cases the condition was chosen as 'identical' but several alternatives are given for each question. For example, in question five the students are asked to find the value of X and five alternatives of the correct answer are given as: 0\ZERO\X=0\X=ZERO\NOUGHT

It can happen that the student will supply a correct alternative which has not been included in the lesson programme. This will be detected by the instructor as he monitors all the student responses. However this answer will not be included with the correct responses in the analysis of each question. This was one of the factors considered when an acceptance level of 70% was decided upon, as discussed in section 3.3.2. The lesson programmes for the six lectures produced for this study appear in Appendix IV.

5.3.2 The Reports

.ce the lesson programme has been entered into the system the lecture may commence. The first instructional sequence is shown followed by a basic question related to this instruction. The students enter their responses into the system and the lecture continues. At any stage during the lecture the instructor may monitor the students' performances via a battery of reports which can be called for and are produced immediately. These reports will be briefly discussed:

(a) The students' responses: Each response made by the students is displayed on the VDU. The instructor may record these responses for later reference.

(b) Detail report: This report indicates which questions have been answered by each student, in addition to revealing whether each answer was correct or incorrect. This report also indicates whether a student has left a particular question unanswered or if he has made more than one attempt at answering it.

(c) Student repo:t: This report would usually be generated at the end of the lecture. It gives the overall results of each student for all the questions. The information it supplies includes the number of questions attempted by the student, the number of correct responses at the first attempt, the final number of correct responses and also the number of correct responses out of the total number of questions attempted.

The above three reports can be very useful to the instructor both during and after the lecture. For example, those students who are experiencing difficulty are easily identified and can be helped during or immediately after the lecture. If the lecture is televised, a poor or ineffective instructional sequence is immediately identified; the television lecture could then be interrupted by the instructor in an attempt to clarify this weak area.

Although it is admitted that the above represent strategies which should be adopted when the sort of information supplied by the reports is available, none of these strategies were employed during the experimentation period. The reason for this was that the object of the exercise was to evaluate the effectiveness of a pre-recorded television programme. If any interruption was made during the showing of the lecture, an uncontrolled variable would have been introduced in the form of the 'live' instructor. Therefore, although the situation was not always entirely satisfactory, no interventions could be made without violating the conditions laid down in the experimental design. The reports referred to were generated during the television lectures, but were used for observational purposes only. Examples of these three reports are given in appendix IV.

A fourth type of report is generated by the system which was essential for the evaluation of the lectures.

(d) Question Report: This report gives a final analysis of how well or badly the individual questions were answered. An example of such a report is given in table 2. The first column (QN) gives the question numbers, in this case from question one (01) to question ten (10). The

Table 2: A Question Report

QN	L%	R%	A%	M%	ANSW.
01	60	60	100	60	2\
02	80	80	100	80	6\
03	60	60	100	60	52\
04	80	80	100	80	X=4\
05	40	40	100	40	0\
06	100	100	100	100	3/8\
07	40	40	100	40	1/16\
08	80	80	100	80	1/8\
09	40	40	80	50	DISCRETE\
10	60	60	100	60	NO

second column gives the percentage of students who finally answered the question correctly (L%). The third column (R%) gives the percentage of students who answered the question correctly at the first attempt. Notice that the figures in these two columns are identical. This is because the students' immediate responses to the questions were recorded and they were not permitted to change any answers at a later stage. The experimental design in this study required that the students' immediate responses were always recorded. The reason for this will be explained in chapter 6 when the design is discussed. The fourth column indicates the percentage of students who attempted each question (A%) and the final column (M%) indicates the percentage of students, out of those who answered the question, who obtained the correct answer. This is the percentage (M%) which was used as a measure of the effectiveness of each instructional sequence.

The M% value was preferred to the R% value for the following reason. After each question was asked a fixed amount of time was set aside for the students to respond. Thus the slower student might not have had enough time in which to respond, although he had found the instruction adequate. However if we take the students who did respond and determine how many of them responded incorrectly, then this figure will be strongly related to the effectiveness of the instruction in achieving its stated objective.

Admittedly, it could be argued that this figure ignores the students who found the instruction so poor that they are totally confused and are unable to offer any answer at all. Although this probability does exist, it is not deemed to have a great affect as the R% and M% values were very similar.

It is also admitted that the situation where the slower learner is being penalized by a fixed-time period is undesirous. However, as was

pointed out at the beginning of this study, the object was to investigate the effectiveness of instruction within the status quo. Many studies have been conducted which describe the advantages of self-paced learning, and these are not disputed here. The pragmatic point of view is that the lecture as a vehicle for teaching is going to be employed at the tertiary level in this country for quite some time and to quite an extensive degree (often though perhaps in conjunction with some form of self-paced learning, as is the case in the course within which this study was conducted). Under these circumstances, how does one best optimize the instructional effectiveness of the lecture? It is from this point of view that the fixed-time period which the students are given to respond in is accepted, but not necessarily condoned. This feature itself is worthy of further research and the Cybercom system could provide a means of detecting the length of time which different students need in which to respond. There is some evidence that a very small change in response speed is indicative of the strength of learning, and this change can be detected by instrumentation (Judd and Glaser, 1969).

The processes which have been described represent the operating mode in which the system was used to enable immediate evaluation of each instructional sequence. It must be remembered that this was a very limited use of the systems potential. One other process was used which is characteristic of the system and will be mentioned briefly.

5.3.3 Immediate Feedback

In addition to monitoring student responses, the Cybercom system enables communication to take place between the instructor and the individual students. A simplified form of this communication facility was used during the television lectures.

At the end of each instructional sequence, immediately after the question had been asked, the instructor communicated with the students by inserting the word 'ACT' into the viewing window of the student panels and causing the two lights to flash slowly. This is achieved by depressing a special function key on the instructor's keyboard. The students are then given a fixed time in which to enter their responses. At the end of this time period, immediately before the next instructional sequence begins, the instructor communicates with the students again by depressing a reinforcement mode button on his console. If the student entered a correct response, the lights in the viewing window flash on and off at a very high frequency; if not, the lights remain off.

There appears to be a certain amount of evidence that suggests that immediate feedback after a student has made a response facilitates learning. Levie and Dickie (1973) made the following points in a review of research findings in this area. Most researchers have found that knowledge of results is of little value following correct responses, casting doubt on the proposition that the confirmation of correct responses has rewarding properties and acts as a reinforcer. Feedback may aid when the learner has responded correctly but is not certain about the correctness of his response. Knowledge of results does facilitate learning when it follows wrong responses. Such feedback permits the learner to correct his mistakes and will diminish the likelihood that he will recall wrong responses as being correct. Feedback which provides the correct answer is better than merely telling the learner he was wrong. Researchers have generally found that knowledge of results is most effective when provided immediately following the response.

When the television lectures were first analysed, an experiment was conducted in an attempt to measure the effect of the immediate feedback. An achievement type test was conducted at the conclusion of the lectures

(a pre-test had been administered before the lectures were shown) which was based solely on material covered in the lectures. Two groups of students were shown the lectures. The first group used the Cybercom system and therefore were given immediate feedback to their responses in addition to viewing the correct answers at the end of each lecture. The second group were shown the lectures without the Cybercom system and therefore received only delayed feedback when they viewed the answers at the end of each lecture. An analysis of variance showed no significant difference on the 5% significance level although the group which had immediate feedback had shown a greater improvement. However it is felt that some independent variables had not been controlled strictly enough in order to glean very much from the experiment (Young, 1978(2)).

For the purposes of this study, one particular mode of the Cybercom system was made use of. However, as it is a multimode communication system there are a host of possibilities for which the system could be used in research involved with teaching and learning. At this point in time these possibilities are still largely unexplored and very little research has been conducted in conjunction with this system.

PART II :

THE EXPERIMENT

6. THE DESIGN OF THE EXPERIMENT

The area which is researched in this study is not classified as a comparison of one type of instruction with another but rather an evaluation of instructional effectiveness and whether instructional effectiveness can be improved once this evaluative feedback has been provided, under certain prevailing conditions.

Any attempt to evaluate a teacher's effectiveness will be confronted with the question: 'What constitutes good teaching?'. Cohen, Rose and Trent (1973) claim that there is no universal set of expectations which define a teacher's role. Instead, an unknown number of sets are held by people from different cultural and socio-economic backgrounds. In order to determine how a particular observer or even a set of observers will judge a teacher, it must first be established what the observers' expectations are with respect to the teacher role. Elton (1975) wonders whether it is possible to assess teaching or indeed if it is possible to assess anything for which there are no absolute criteria.

It is not intended to become involved in a debate of this nature and therefore this chapter will begin with a delimitiation of the criteria which were used to evaluate the instruction and the conditions under which the instruction took place. The hypotheses to be tested will then be stated and the identification and control of the variables expounded. Finally, a description of the sampling process adopted in the selection of the students will be given.

6.1 The Purpose of the Experiment

6.1.1 Some Prevailing Conditions

The motivation for this study arose from the teaching conditions which were prevalent in an introductory statistics course which is administered to large numbers of students. In an attempt to improve these conditions the use of media was investigated. Certain aspects of the systems approach to instruction were used in the selection of media which included the utilization context (the ease with which it might be administered to large numbers) and evaluation procedures (the ease with which developmental evaluation and revision might be accomplished). (Levie and Dickie, 1973). Other attributes have to be considered in media selection and Glaser and Cooley (1973) state that it is 'platitudinous' to say that instructional systems need to adapt to the requirements of the individual learner, but how to bring this about in educational institutions which are overwhelmingly geared to mass group instruction is a fundamental question.

After consideration of these and other conditions, it was decided to use the medium of television for the lecture situation in the introductory statistics course. This is only one of several components which constitute all the teaching situations associated with the course. Although this decision may be construed as the furtherance of mass education, one factor which was instrumental in this decision was the object of releasing lecturers from routine, repetitive situations and increasing their availability for teaching situations which were more individualised and interactive than the formal or traditional lecture.

This investigation was initiated within this prevailing framework. Given these conditions and the fact that televised instruction would be used in a fixed-time lecture period with large numbers of students in

attendence, how could these lectures be evaluated and subsequently improved?

Several other prints with regard to the prevailing conditions are worth noting; these are probably not unique to this University but are applicable to many institutions in higher education. Different lecturers are responsible for the presentation of different sections of the course; the section is usually the area in which the lecturer is academically regarded as a specialist. Not all lecturers are, although it would seem that they are expected to be, specialists in education and teaching. It is rather ironic that academics and scholars who set such great store by the analytical powers of the mind so rarely apply those powers when considering their own teaching (University Teaching Methods Unit, 1978). If the lectures are to be evaluated where the underlying motive is directed towards their improvement, a method must be adopted which is acceptable to all the lecturers and which is effective.

The television lectures produced for this study were constructed where these conditions and circumstances were dominant factors.

6.1.2 The Model to be Tested

A topic in the introductory statistics syllabus was chosen which was entirely new to the students. In other words it had never been included in the secondary school mathematics syllabus and the probability of any students having covered it previously was minimal. This was to ensure that any evaluation which would take place during a lecture would be based on that instruction and not on any previous teaching. The topic chosen was 'An Introduction to Probability Distributions' and included the consideration of some properties of probability distributions in addition to the examination of two particular probability distributions. Several mathematical skills and a certain amount of the mathematical background required for these lectures should have been previously

acquired by the students. As most lecturers assume that students have acquired the necessary background and skills from previous work when new subject matter is being dealt with, it was assumed that the students had managed this acquisition when the topic was approached. This is very seldom the true state of affairs and was blatantly pointed out when the lectures were evaluated, as is described in chapter eight.

This topic would normally be covered in about six forty-five minu lecture periods by a 'live' lecturer. Therefore six television lectures, each of an equivalent length to that of the 'live' lectures, were produced. The content and presentation of the lectures were based, to as great an extent as possible, on the same lines as that of the 'live' lecture. The intent here was twofold. First, it was accepted that if a team of teaching and subject experts had composed and produced the lectures, they would have been of a far higher quality. However, that type of model would not be acceptable to the 'average' lecturer who usually prefers to present his lecture with his own style and personality. Many lecturer's do not have the time (or are not prepared to make the time) to spend with such a team examining and improving their teaching. Secondly, many lecturers do not have the background and training required to understand and apply many teaching techniques and learning principles. In short, they are not properly qualified educationists and are not likely to become so and therefore would not be likely to adopt a lecturing method which imposed these conditions on them. It has already been shown (Gropper, 1967) that 'programmed' television instruction can be improved by simply adopting tried and tested principles of programmed instruction.

Thus the changes imposed on the transition from a 'live' to a televised lecture were kept to a minimum; the intent being that these changes could be outlined to any lecturer in a short amount of time. The changes that were imposed can be described as follows:

(1) A general aim or intention was stated for each lecture: a statement which outlined in general terms what the students should have achieved by attending the lecture.

(2) The specific abilities which students should achieve at various stages during the lecture were described in terms of observable student behaviour. In other words, what should the students learn from each instructional sequence in the lecture, and how will they demonstrate that they have learnt what it is that they were supposed to learn?

(3) A very simple or basic question (or two) was generated for each instructional sequence which tested the achievement of the abilities stated in (2). These questions were inserted at the end of each instructional sequence and were displayed for a fixed period of time.

(4) Television graphics were prepared before a lecture was recorded, although occasional notes were still made on the chalkboard during a lecture. In order to assist the student, a green background was used with certain graphics to indicate expressions, formulae or statements which were worth taking note of; in all other cases a blue background was used.

The content and presentation was then left to the lecturer himself. Only slight modifications were made to his 'normal' lecture so as to include some of the general principles of criterion-referenced instruction like writing objectives in behavioural terms. Admittedly, some of the learning principles outlined in chapter two now form part of the lecture whereas this might not have been the case previously. For example, when students answered the questions during the lecture, they were obtaining appropriate practice of the skills they were supposed to be learning; when they were given feedback via the Cybercom system the principle of knowledge of results was being applied in an optional sense.

The central concern of this study however was not whether the lecturer's instruction had now improved from what it was, although this

might well have been the case. The model described here is one which any lecturer could, relatively painlessly, adopt and which would represent a minor transformation from the 'standard' lecture which he usually delivers.

6.1.3 The Hypotheses

The procedures adopted in the evaluation of this instructional model were based on the following two null hypotheses:

1. During the presentation of recorded television lectures on introductory statistics, students' attainment of pre-stated behavioural objectives cannot be assessed via a computer based teaching system.

2. Students' attainment of pre-stated behavioural objectives during the presentation of recorded television lectures on introductory statistics cannot be improved by analysing the effectiveness of the television lectures and administering changes to parts of the television lectures.

The experimentation necessary to test these hypotheses leads to the postulation of a third hypothesis which was investigated:

3. There is no significant difference between the students' attainment of pre-stated behavioural objectives during the presentation of recorded television lectures on introductory statistics and the attainment anticipated by the presenter of the television lecture.

Several variables had to be identified and controlled before any results were to be obtained which could be used to test the null hypotheses.

6.2 The Identification and Control of the Variables

6.2.1 The Criteria for Evaluation

As was mentioned in the introduction to this chapter no absolute criteria exist or are universally accepted when instruction is to be evaluated. Therefore the criteria which were to be used to evaluate instruction in this study must be clearly stated. The television instruction was judged using the criterion of learner growth, where learner growth was assessed in terms of the attainment of well-defined behavioural objectives. A full discussion of the type of behavioural objectives used in the television instruction was given in section 3.3.

The development of behavioural objectives as an attempt to deal with the very basic question of what is being evaluated when assessing teacher competence is a relatively new approach. Performance tests of teaching proficiency have been developed where teachers are given sets of explicit instructional objectives and asked to teach specifically to them; instructional effectiveness is then assessed in terms of the teacher's ability to produce the student behaviour changes prescribed by the objectives (Roueche and Pitman, 1972). Although several aspects of this approach are often subject to criticism, as outlined in section 3.1.2, this is the type of approach which was used to evaluate the effectiveness of the television instruction.

An important aspect of this approach is the means that are used in order to determine if these objectives have been achieved. The first consideration is that the objectives must be stated in terms of observable student behaviour. Secondly, the behaviours which the students are asked to per-

form (the questions which they are asked) must be the behaviours which they are required to attain. Thus the questions which are used to determine the attainment of a particular behaviour must be valid; if they are not, uncertainty will remain as to what it is that is being measured.

Thus the validity of the questions which were used is an important variable in this experiment and one which must be strictly controlled throughout. The validation of the questions is fully discussed in the next chapter.

6.2.2 The Instructor Effect

In order to ensure that any behavioural changes be attributed to the television instruction only, no interventions or interruptions were made during the replaying of a television lecture. In addition, no comments were made prior to or after a replay. In fact, the only motive for the presence of a 'live' lecturer during the replays was that the Cybercom system had to be operated and to enable the lecturer to monitor the responses as they were being made.

The students were informed that the purpose of recording their responses was to evaluate the effectiveness of the instruction and not to evaluate them.

The same lecturer was responsible for the presentation of all six lectures which covered the chosen topic, in order to prevent any instructor interaction which might have occurred if more than one lecturer had been used.

6.2.3 The Response Delay

In the majority of methodological experiments in education, some type of measurement is made immediately after the administration of a treatment. Often, a delayed measurement is made at some later stage. However, the longer the delay before the measurement is made, the greater the probability that some uncontrolled variables will interact with the treatment. Thus the initial measurement is made as soon as possible after the treatment.

In many experiments, this is often a difficult exercise and interactions which occur with the treatment are usually difficult to control. A central feature of this experiment was the immediate measurement which took place after each instructional sequence which was achieved by using the Cybercom system. Thus any interactions which might occur during a lecture were kept to an absolute minimum.

6.2.4 The Sampling Procedure

Three samples of students were used in order to evaluate the television lectures. The sampling process used is an important feature of any experiment and the one adopted will be fully described in the next section.

6.3 The Student Samples

6.3.1 The Measurement Variables

The selection process employed in obtaining the student samples was one where students were randomly selected from the parent student population according to the first letter of their surnames. This did not ensure that the samples contained students of identical mathematical ability as would have been the case if matched pairs had been selected. The measurement variables were designed in such a way as to obviate this necessity.

Two measurement variables were of importance in order to test the hypotheses. In order to test the first hypothesis, an attempt was made to measure the attainment of pre-stated behavioural objectives during a television lecture with the aid of the Cybercom system. The actual level of attainment was not important to this part of the experiment.

However, this initial exercise in evaluating the instruction acted as a guide as to which sequences in the lectures were unsatisfactory and should be changed. Thus it was important that the sample for this evaluation was representative of the 'population' of students. The acceptance

level of instructional effectiveness had been set at 70%. The initial sample was therefore used as an indicator of the achievement of this level for each instructional sequence. In order to ensure that this indication was not significantly different from that which would have occured for the population, a statistical test was conducted on the results of a pre-test which was administered to all the students at the commencement of the statistics course. The results of this pre-test are given in section 6.3.2.

The measurement variable which was used to test the second hypothesis was the differences between the scores obtained by two samples of students for the instructional sequences which were left unchanged and for those that were changed. These two samples were not perfectly matched and their abilities not identical. This requirement is not necessary as the differences between the scores of the two samples were being compared for two sets of instructional sequences: the changed and unchanged instruction.

This point can be illustrated graphically as follows: if the two groups were identical and the instruction was also identical (this is the instruction which was left unchanged), then the differences between scores for each question which was asked at the end of every instructional sequence, would be expected to be distributed in a normal manner about the value of 0, as shown by distribution (1) in figure 12. If the instructional sequences which were changed caused a difference in effectiveness (either positive or negative) then this change in effectiveness would influence the results obtained by the group who was subjected to the changed instruction (the other group received the original instruction) and the differences in their scores would no longer be distributed about 0, but some value, x_1 as indicated by distribution (2) in figure 12. These two distributions could then be tested for any significant differences.

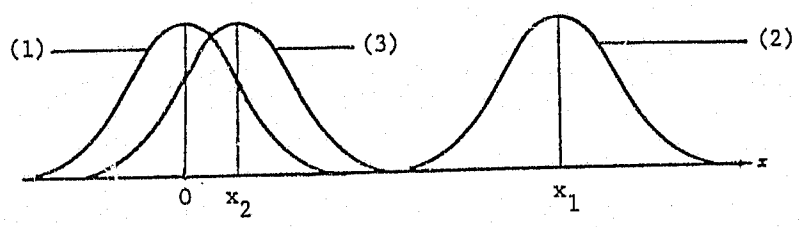


Figure 12: Expected Distributions for Identical Groups.

X: Differences between scores for questions related to,

(1) Unchanged Instruction,

(2) Changed Instruction, where a difference is indicated,

(3) Changed Instruction, where no difference is indicated.

If the changed instruction produced no effective change on the students, then the results should be very similar to those obtained by the group who were given the original instruction and the differences be distributed about 0 or a value very close to 0, say x_2, as indicated by distribution (3) in figure 12.

If, however, the two groups were not identical, then there would be a difference in their results for the identical instruction. These differences would not be distributed about 0, but some other value, say x_3. This is indicated by distribution (1) in figure 13. A similar argument can now be employed to explain that changed instruction which was effectively different can be represented by distribution (2), about some value x_1, and if itwas not different, itwould be distributed about some value x_2 which is very close to x_3; this is illustrated by distribution (3), figure 13.

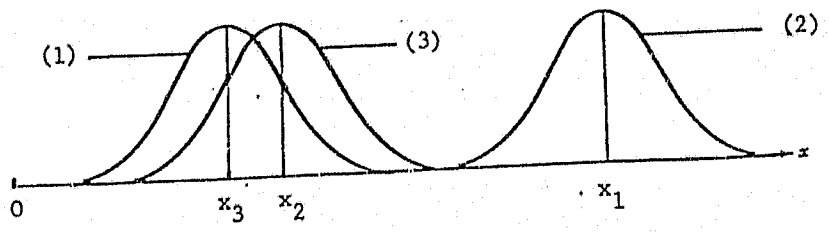


Figure 13: Expected Distributions for Non-Identical Groups.

X: Differences between scores for questions related to,

(1) Unchanged Instruction,

(2) Changed Instruction, where a difference is indicated,

(3) Changed Instruction, where no difference is indicated.

It can be argued that two vastly different samples could introduce new, uncontrolled variables and that the distributions would not follow the pattern described. In order to obviate this possibility, two similar, but not necessarily identical samples were chosen. These two samples were also similar to the sample which was used for the initial evaluation of the instruction and all three samples were representative of the student 'population.' This was established via the pre-test, the results of which are analysed and described in the next section.

6.3.2 Pre-Test Results

Three samples were chosen from the student 'population'. These samples differed in size. The student 'population' (those taking the introductory statistics course) numbered about 1700. This population was divided into eighteen separate groups for administrative and logistical reasons. Three of these groups constituted the three samples used in the experiment. It

was decided not to use samples specially constituted for the experiment so as to avoid any Hawthorne effect[1]. It was however, desirous that these three samples were not significantly different in terms of mathematical ability as indicated in section 6.3.1. A pre-test was administered to all the students in order to determine their entry skills at the commencement of the course. A copy of the test is given in appendix V.

The first analysis concerns the homogeneity of the three samples. The null hypothesis to be tested is: there is no significant difference between the three samples in terms of the pre-test results.

---

1. 'In many experiments there is a danger that subjects will behave artificially. Volunteers for an experiment know that it is not for real and may unconsciously fall into certain poses. Cast into a let's - pretend role they often respond according to their perception of what is socially normal. Or subjects may baulk at being manipulated and try to outthink the experimenter. They may be hypercooperative or hypercompetitive with their fellow subjects. The most famous example of a laboratory effect is the Hawthorne study, in which selected factory workers were so pleased at the experimental attentions of the management that they produced steadily more with each new change, regardless of whether it was an improvement. The highest productivity came with the final change, which took away all breaks, hot lunches, and incentive payments. Needless to say, generalization to all factory workers was rather unsafe. This form of guinea-pig reaction, where subjects respond much more to the fact of being subjects than to the treatment, has been dubbed the Hawthorne effect.' Ross, J. and Smith, P. Orthodox experimental designs. In Blalock, H.M. and Blalock, A.B. (Eds.). Methodology in Social Research. McGraw-Hill, 1968, p.34[illegible].

A one-way analysis of variance was conducted. The results of this analysis are given in table 3. The 95% confidence limit was chosen; thus differences were sought at the 0,05 significance level (S.L.). The F value at the 0,05 significance level for $\nu_1 = 2$ and $\nu_2 = 120$ degrees of freedom is 3,07 and for $\nu_1 = 2$ and $\nu_2 = \infty$ degrees is 3,00 (Pearson, 1966). Thus the F value for $\nu_1 = 2$ and $\nu_2 = 184$ degrees of freedom, as is the case in the analysis given in table 2, will lie between 3,00 and 3,07. The F value in the analysis is 2,305. Therefore the null hypothesis of no significant difference can be accepted.

The second analysis concerns the representativeness of the three samples with respect to the student population. Three separate tests were conducted, one for each sample. The null hypothesis tested in each case was:

There is no significant difference between the sample and the population in terms of the pre-test results.

As each sample had a size greater than thirty, a t-test is unnecessary and the analysis can be conducted in terms of a normal approximation (Chao, 1974). A significance level of 0,05 was again chosen. The population mean (μ_o) was equal to 34,299 and the population standard deviation (σ) was 19,211. The results of these analyses are given in table 4.

As the z-scores for all three samples were within the 95% confidence limits, the null hypothesis was accepted in each case.

Thus before evaluation of the television instruction took place it was established that there were no significant mathematical differences between the three samples in terms of the pre-test and that the samples were representative of the student population. One other point that had to be established before any evaluation took place was that the questions that were asked during the lectures were related to the stated objectives. A description of this validation process is given in the next chapter.

SPSS ANALYSIS OF PRE-TEST

- - - -

SPSS ANALYSIS OF PRE-TEST

- O N E W A Y -

VARIABLE MARK

ANALYSIS OF VARIANCE

SOURCE	D.F.	SUM OF SQUARES	MEAN SQUARES	F RATIO	F PROB.
BETWEEN GROUPS	2	1593.6650	796.8323	2.305	0.1027
WITHIN GROUPS	184	63621.8828	345.7710		
TOTAL	186	65215.5469			

GROUP	COUNT	MEAN	STANDARD DEVIATION	STANDARD ERROR	MINIMUM	MAXIMUM	95 PCT CONF INT FOR MEAN		
GRP01	99	34.8687	18.5701	1.8664	1.0000	75.0000	31.1649	TO	38.5724
GRP02	34	30.2059	17.6116	3.0204	4.0000	65.0000	24.0609	TO	36.3508
GRP03	54	38.8889	19.2262	2.6164	11.0000	83.0000	33.6411	TO	44.1366
TOTAL	187	35.1818	18.7249	1.3693	1.0000	83.0000	32.4805	TO	37.8832

Table 3: One way analysis of variance of the pre-test results for Groups 1,2 and 3.

Table 4: Significance Tests for Three Samples

$$\mu_o = 34,299; \quad \sigma = 19,211.$$

$H_o : \mu = \mu_o$

$H_1 : \mu \neq \mu_o$

At the 0,05 significance level, for a 2-tailed test: $-1,96 \leq z \leq 1,96$ if the null hypothesis (H_o) is to be accepted.

Sample 1: $\bar{x} = 34,87; \quad n = 99.$

$$z = \frac{\bar{x}-\mu_o}{\sigma_{\bar{x}}}$$

$$= \frac{\bar{x}-\mu_o}{\frac{\sigma}{\sqrt{n}}}$$

$$= (34,87-34,299)/(19,211 \times \frac{1}{\sqrt{99}})$$

$$= 0,296.$$

Sample 2: $\bar{x} = 30,21; \quad n = 34$

$$z = \frac{\bar{x}-\mu_o}{\sigma_{\bar{x}}}$$

$$= \frac{\bar{x}-\mu_o}{\frac{\sigma}{\sqrt{n}}}$$

$$= (30,21-34,299)/(19,211 \times \frac{1}{\sqrt{34}})$$

$$= -1,241.$$

Sample 3: $\bar{x} = 38,89; \quad n = 54$

$$z = \frac{\bar{x}-\mu_o}{\sigma_{\bar{x}}}$$

$$= \frac{\bar{x}-\mu_o}{\frac{\sigma}{\sqrt{n}}}$$

$$= (38,89-34,299/(19,211 \times \frac{1}{\sqrt{54}})$$

$$= 1,756.$$

7. VALIDATION OF THE EVALUATIVE MEANS

The effectiveness of the television instruction was to be judged in terms of student achievement of pre-stated behavioural objectives. This achievement was measured in terms of student responses to questions which elicited these behaviours. Some validation technique had to be used in order to ensure that the questions asked did in fact elicit the required behaviours. In short, the questions had to measure what they were supposed to measure.

Measurement practices employed in the construction and analysis of tests have undergone major developments and refinements in many psychological and educational studies and it is not intended to attempt even a summary of this vast field. Some central ideas in this field, such as test validity and reliability will be outlined in that they are relevant to the present situation. This chapter will outline the adaption of these ideas to the validation of the questions used in the lectures and the technique and result of this process will be described.

7.1 Validation Procedures

7.1.1 Reliability and Validity

Two important attributes of a test are its reliability and validity. There are several operational definitions of the reliability of a test and in almost all of them high reliability means that repeated measurements will give nearly the same results (Nedelsky, 1965) Thus the reliability of a test can be considered to be a measure of its consistency. A closely related concept is that of test validity. The validity of a test is the extent to which a test measures what it purports to measure. The relationship between these

two concepts is such that a valid test will also be reliable whereas a test which is highly reliable (gives consistent results) is not necessarily valid. If a test is not reliable, the inconsistency of its results will immediately render it invalid.

Several methods of determining the reliability of a test have been developed, but will not be discussed here. An attempt to bring a set of similar expectations and common language to the validity arena was made by French and Michael (1966). This led in part to the development of three types of validity measures in educational and psychological measurement. These three approaches can be summarized as follows:

1. Content validity: The adequacy with which the content of the test samples the behaviour or content domain about which inferences are to be made. This is usually determined by inspection and this approach relies heavily on human judgement.
2. Criterion-related validity: The correlation of the performance with some external criterion. An example of this would be the correlation between an IQ test and some later test measures.
3. Construct validity: A hypothetical construct presumed to account for test performance is identified; one or more hypotheses are established regarding test performance based on this construct; the hypotheses are then tested by empirical methods.

The content validity approach was chosen for the validation of the questions used in the television lectures. The suitability of this approach for criterion-referenced questions will be outlined in the following sections.

7.1.2 A Priori and A Posteriori Approaches

Perhaps the most common statistical analysis used in improving a test's reliability (and if the major objectives are properly represented in the test, also its validity) is item analysis, in which the correlation between

the test as a whole and an individual item is calculated (Nedelsky, 1965). A short cut method of calculating this correlation is usually adopted and leads to an index of discrimination. This index is obtained by using the total test scores to separate the students into 'high' scoring and 'low' scoring groups. If the high group does much better than the low group on an individual test item, then that item is said to show good or positive discrimination. If the high and low groups do equally well on an item, it has no discrimination and if the low group does better than the high one, it has negative discrimination.

Thus the validity and reliability of a test are usually improved by aiming at achieving items with positive discrimination indices. As this is an empirical method, the approach to improved test reliability and validity is mainly an a posteriori approach. An a priori approach is usually adopted to the extent that test writers always try to write good tests, but this is mostly a subjective attempt. If a discrimination index is to be used to improve items then data must be collected on the items' performance and test improvement is effected via a posteriori methods.

7.1.3 A Priori Validation of Criterion-Referenced Tests

The remarks made so far in this chapter concerning test analysis apply to what was defined in chapter three as a norm-referenced test. The questions in the television lectures were based on a criterion-referenced measurement approach. If the methods for improving test validity and reliability were imported en masse and applied to criterion-referenced tests, several difficulties could result.

The point was made in chapter three that one of the basic requirements of a norm-referenced achievement test is that it should spread students out so that subtle comparisons can be made among them. A test item that is most effective in spreading examinees out is one with a positive discrimination index: it makes the high group's score even higher and the low group's

score lower and thereby increases the spread in the final result. Items which do not in any way affect the spread of the results are those with a zero correlation index. These items include questions which are answered equally well by both the high and low groups and they are the items which are usually discarded or changed when a test is revised. Thus items which are answered correctly by large proportions of students are excised from the test. For purposes of spreading examinees out, these items are redundant. Popham (1978), in his introductory text on criterion-referenced measurement, claims that this is one of the factors which makes some of the techniques of test validation and reliability used in norm-referenced testing, inappropriate for criterion-referenced tests.

He argues that test items on which students perform particularly well tend to be items covering the very concepts that instructors thought important enough to stress. This leads to the situation where the more important the topic is, the more likely a teacher is to emphasize it by devoting instructional time to its mastery; the more instructional time devoted to a topic, the more likely that the norm-referenced test items related to that topic will be answered correctly by many examinees; the more often a test item is answered correctly by many examinees, the more likely that item is to be removed from that test. Thus a norm-referenced test which has often been revised and is highly reliable and valid will tend to omit the most important and often taught areas. Tyler (1973) cites a series of international education studies in which the only items on achievement tests that discriminated among pupils were based on things not taught in schools; the items covering school related topics had long since been jettisoned from the tests.

This is possibly an extreme view but nevertheless does point out some of the difficulties which would be encountered if this technique was used to improve the validity and reliability of a criterion-referenced test.

The items with zero discrimination (those answered correctly by most or none of the students) are precisely the ones required in a criterion-referenced test in order to indicate the attainment or otherwise of a required objective. These factors were instrumental in the decision to use the content validity approach when validating the lecture questions. This is an a priori approach which relies on judgemental techniques as opposed to the empirical methods which are more prevalent in a norm-referenced test validation procedure.

7.1.4 The Validation Techniques

There is substantial support for the view that subject content specialists are best qualified to validate items or questions when using a content validity approach. Coulson and Hambleton (1974) conclude that content specialists' ratings, along with empirical procedures, provide an excellent basis for establishing content validity of domain-referenced test items. In a later study Rovenelli and Hambleton (1976) report favourably on the use of content specialists in judging the quality of criterion-referenced test items. They indicate that there is considerable evidence to suggest that content specialists can complete their ratings quickly and with a high degree of reliability and validity.

This is the point of view adopted in this study and a team of five content specialists was used to validate the questions inserted into the television instruction. Popham (1978) claims that in addition to using content specialists, a method of systematic human judgement must be employed. This systematic method would usually include the following steps:

(a) The services of a reasonable number of content specialists (preferably more than three) must be secured.

(b) A list of the competencies or objectives to be tested must be read by the content specialists.

(c) The content specialists must then match the questions or test items to the appropriate objective or competency being measured.
(d) The questions should be arranged in a random order and a one to one correspondence should not always apply. In other words, more than one question (or perhaps none) could apply to a single objective and vice versa.
(e) If a majority of the content specialists agree that an item does not match its intended objective, then that item should be discarded or reviewed.

This is the validation technique which was employed to ensure that the questions used in the television lectures were indeed valid. The implementation of the technique is discussed in the next section.

7.2 The Matching of Questions and Objectives

The instructional model to be evaluated which was described in chapter six required that, after a minimum number of modifications had been made to the lecture in its 'traditional' form to accommodate the transition from a 'live' to a television presentation, a general objective or aim for each lecture be stated. Secondly, specific behavioural objectives had to be stated for each instructional sequence and thirdly, questions had to be generated which would test the attainment of these objectives. These three phases will be described in this section.

7.2.1 The General Objectives or Aims

The titles of the six lectures which were evaluated, together with the general objective or aim of each were:

Lecture 1

M_{35}: Probability Distributions

Aim: The students will understand the meaning of a probability distribution and the concepts related to probability distributions.

Lecture 2

M_{36}: Expected value of a Random Variable

Aim: The students will grasp the significance of the expected value of a random variable.

Lecture 3

M_{37}: The Variance of a Random Variable

Aim: The students will grasp the significance of the variance of a random variable.

Lecture 4

M_{38}: The Binomial Probability Distribution

Aim: The students will understand the Bernoulli and Binomial probability distributions.

Lecture 5

M_{39}: Properties of the Binomial Distribution

Aim: The students will know the properties of the Binomial distribution.

Lecture 6

M_{40}: The Poisson Probability Distribution

Aim: The students will understand the Poisson probability distribution.

The operative words in all these aims are all open to various interpretations. To 'understand' the Poisson distribution will probably mean different things to different instructors. Thus the aims are not explicit enough to be useful in terms of measuring their achievement. In order for this to be done, the aim must be more specific in describing what the student will be doing to demonstrate the achievement of this aim.

Once the aims are stated in terms of observable student behaviour, they are usually referred to as behavioural objectives. In order to evaluate the effectiveness of the instruction in enabling the students to achieve these objectives, the model used in this study specified that the objectives must be behaviourally stated. The next step therefore was to state the objectives for each lecture in behavioural terms.

7.2.2 Specific Objectives

More than one specific objective was required for each lecture. When moving from a general lecture aim to behaviourally stated objectives it was found that the students were expected to attain several behaviours in each lecture. Thus each of the six lectures had more than one behaviourally stated objective. This in effect reduced each lecture to a series of instructional sequences, where each instructional sequence had a specific objective.

The final step in the preparation of the lectures was the generation of basic questions which were designed to test the achievement of each objective. It was often found that several questions could be generated for each instructional objective which tested the formative steps that had to be assimilated before the entire behaviour was attained. Therefore more than one question was often generated for each sequence and inserted either at different stages of the sequence or one after the other at the end of the sequence.

Appendix I gives the behaviourally stated objectives followed by the questions generated for each lecture. The questions are given in the order in which they occur in the lectures. The objectives are not given in the order in which the students are to achieve them. This 'achievement' order will become apparent when discussing the matching results in section 7.3.

7.2.3 Matching Process

The reason for the objectives not being in the order in which they weresupposed to be achieved is that Appendix I is a copy of the list of questions and objectives which was handed to the content specialists for matching. Therefore the objectives were placed in a random order so as to eliminate any patterns which the content specialists may have detected while validating the questions. The response form which was given to each content specialist is included at the start of Appendix I.

Five content specialists were used. They were all University lecturers with substantial experience in teaching the introductory statistics course. Each content specialist was given a list of the behavioural objectives and questions for each of the six lectures. They were requested to read through the objectives and questions for each lecture and then mark down next to each objective, on the response form provided, the number of the question or questions which matched or related to a particular objective. They were informed that more than one question could be related to one objective and some questions might not be related to any objective. No time limit was set for the completion of this matching process.

7.3 Results

7.3.1 The Validity Ratings

No additional questions were placed on the list given to the content

specialists. Therefore all the questions should have been matched with an objective. Although it was not intended, one question could have been related to two or more objectives. This situation could arise when the attainment of an objective during the initial stages of a lecture was a prerequisite for the attainment of subsequent objectives. In this instance, the question which tested the attainment of the initial objective could be interpreted as being related to the subsequent objectives.

Although this occurrence was not indicated explicitly as a possibility to the content specialists, they all recorded instances of this type of matching. It was decided that a question would be considered to be matched to an objective if a majority of content specialists indicated this as being the case. However, questions which received a matching from only three out of the five specialists should be closely scrutinized and possibly 'massaged' to improve their validity rating.

A summary of the matchings recorded by the five content specialists for the six lectures is given in table 4. Lecture 4 contains eight questions and Lecture 5 nine. The other lectures all contain ten questions. The second column indicates which objective was matched to each question by the content specialists. The order in which the objectives were to be attained by the students now becomes apparent. For example, the order of attainment for Lecture 1 is B; C; A. If the objective column for Lecture 1 in table 4 is scrutinized, this is the order in which the objectives occur.

An exception occurs in Lecture 3 where the attainment order of objectives A and B appear to alternate. The objective order for questions 2 to 7 is given as: B; A; B; A; B; B. This apparent exception is easily explained if objectives A and B are more closely examined. Objective B can be considered to be a dual objective. It required that students are able to calculate the variance of:

Table 5: Summary of matchings recorded by five content specialists

Lecture 1 (H_{35})

Question	Objective	Validity	Duality	No Match
1	B	5		
2	B	5		
3	B	5		
4	B	4		
5	B	4		
6	C	5	B(3)	
7	C	5		
8	C	4		1
9	A	5		
10	A	5		

Lecture 2 (H_{36})

1	A	4		1
2	A	5		
3	B	5		
4	D	5		
5	C	5		
6	E	5	B(3)	
7	E	5		
8	E	5		
9	E	5		
10	E	5		

Lecture 3 (H_{37})

1	C	4		
2	B	5		
3	A	5		
4	B	5		
5	A	5		
6	B	5	A(3)	
7	B	5		1
8	D	5		
9	D	5		
10	D	5	B(5)	

Lecture 4 (H_{38})

Question	Objective	Validity	Duality	No Match
1	C	4		1
2	C	5		
3	C	4		
4	C	4		
5	A	4		1
6	B	5		
7	B	5		
8	B	5		

Lecture 5 (H_{39})

1	C	5		
2	C	4		
3	A	3		
4	A	3	C(3)	
5	D	5		
6	D	5		
7	B	5		
8	B	5		
9	A	5		
10	A	5		

Lecture 6 (H_{40})

1	E	5		
2	E	5		
3	A	5		
4	D	5		
5	B	5		
6	B	5		
7	C	5		
8	C	5		
9	F	5		

i) a sample, and

ii) a population.

Objective A can also be considered to be a dual objective. It required that students are able to recognise that:

i) σ^2 and

ii) $E(X-\mu)^2$,

are both equivalent terms for the variance of a population, Var(X).

Therefore the mixing of the order of objectives A and B is due to the achievement of parts of these objections. The order of these multi-part objectives is B(i); A(i); B(ii); A(ii); B(ii); B(ii), which explains the pattern for questions 2 to 7 given in Table 4.

The column headed 'Validity' gives the number of content specialists who matched the question to the objective.

If a majority of content specialists indicated that a particular question was related to a second or earlier objective, then this 'dual' question's other objective is given in the 'Duality' column. The actual number of content specialists who made this indication is given in brackets after the objective.

The final column in table 4 is used to note the number of content specialists who did not match a question to any of the stated objectives.

7.3.2 Disucssion

The validity ratings of the five content specialists shows a high correlation with the intended question - objective matchings. The net result was that none of these questions were changed before the television lectures were shown to the first group of students. However, several points which arose from the results given in table 4 should be noted.

The first has already been made: the achievement order of the objectives can be determined by noting the order in which they occur in the 'Objective' column.

Secondly, the validity ratings or matchings can be summarised as follows: All fifty seven questions were correctly matched by a majority of the content specialists; forty four of these questions received a validity rating of 5 (all five content specialists matched the question to the same objective); ten questions received a rating of 4 and only three questions received a rating of 3. These three questions (Lecture 3, question 7 and Lecture 5, questions 3 and 4) were closely scrutinized and it was decided not to change them in any way; at least not until the results of the first group of students were available.

A majority of content specialists indicated that six questions were related to a second objective. An examination of table 4 reveals that this second objective always appears prior to the matched objective. This suggests that the attainment of the matched objective is dependent upon the prior attainment of the second objective. If the actual objectives are consulted in Appendix I, this suggestion is confirmed. All six of these 'dual' questions assume the attainment of a previous objective. This point is also discussed when a description of the analyses of, and the changes to, the television lectures is given in section 8.2.

Finally, four questions were not matched to any of the stated objectives by only one content specialist in each case. This proportion was not considered large enough to warrant any changes to these questions or their objectives.

The questions were now considered to be competent in measuring what they were supposed to measure: the degree of attainment of behaviourally stated objectives.

8. THE RESULTS AND ANALYSES

This chapter focuses on the major part of the experimental work completed in this study. The first stage of the experiment was the initial evaluation of the instructional effectiveness of the television instruction. Secondly, instructional sequences which did not achieve the 70% acceptance level in the initial evaluation were analysed. An attempt was made to find the source of failure in each of these instructional sequences and changes to the instruction were made in order to improve (or attempt to improve) the instruction. Thirdly, the television lectures were shown to two groups of students. The first of these two groups was shown the original instruction; the second was shown the changed instruction which included some of the original instructional sequences and some changed or 'improved' sequences.

The three sections in this chapter describe the implementation of each of these stages and the results which were recorded at the completion of each stage.

8.1 The Preliminary Results

8.1.1 Attainment of the Acceptance Level

The six television lectures were prepared and recorded with a structure as outlined in chapters 4 and 6. The lectures were then replayed to all the students attending the introductory statistics course. One of these groups of students, which will now be referred to as group 1, attended the replays in a lecture theatre which was 'fitted' with the Cybercom computer system. Their responses to the questions in the

lectures were then recorded as described in section 5.3. A transcription of these lectures, together with video-tape reference points, can be found in appendix III.

The results for group 1 are given in table 6. As these questions were all judged as valid (section 7.3) these results also represent the proportion of students who achieved the objective for each instructional sequence. For example, the first five questions in lecture 1 yielded results which were all above the acceptance level of 70%. These five questions were all related to objective B for lecture 1 (see table 5, section 7.3). Thus the instruction which was directed at achieving objective B can be considered to be satisfactory in terms of the specified acceptance level. Questions 6,7 and 8 then yielded results which were far below the level of acceptance. As these three questions were all related to objective C for lecture 1, the instruction associated with this objective was entirely unsatisfactory. Finally, questions 9 and 10 were related to objective A. Question 9 achieved the acceptance level but question 10 did not. Thus the instruction associated with this objective was partially satisfactory and the area related to question 10 should be closely scrutinized and possibly some adjustment made in an attempt to improve its effectiveness. A detailed description of this procedure is given in the next section.

The overall summary of the results in table 6 is disappointing. The acceptance level was attained by twenty three questions; thirty four questions failed to achieve this level. None of the six lectures managed a mean score of 70% or more and the mean score for all the questions was 52%.

Table 6: Group 1 Results for Television Lectures 1-6

Lecture 1 (M_{35})			Lecture 2 (M_{36})			Lecture 3 (M_{37})		
Question	Score	Change	Question	Score	Change	Question	Score	Change
1	72%		1	93%		1	70%	
2	91%		2	81%		2	5%	*
3	74%		3	18%	*	3	74%	
4	94%		4	26%	*	4	24%	*
5	90%		5	31%	*	5	43%	*
6	9%	*	6	43%	*	6	80%	
7	28%	*	7	12%	*	7	19%	*
8	33%	*	8	78%		8	56%	*
9	89%		9	77%		9	14%	*
10	66%	*	10	68%	*	10	10%	*
Mean	65%		Mean	53%		Mean	40%	

Lecture 4 (M_{38})			Lecture 5 (M_{39})			Lecture 6 (M_{40})		
Question	Score	Change	Question	Score	Change	Question	Score	Change
1	79%		1	45%	*	1	42%	*
2	40%	*	2	43%	*	2	81%	
3	41%	*	3	76%		3	91%	
4	48%	*	4	52%	*	4	10%	*
5	72%		5	76%		5	17%	*
6	5%	*	6	98%		6	28%	*
7	4%	*	7	70%		7	65%	*
8	32%	*	8	71%		8	17%	*
			9	88%		9	21%	*
			10	65%	*			
Mean	40%		Mean	68%		Mean	41%	

Mean for 57 Questions: 52%

*Questions which failed to elicit an acceptable response level (70%)

8.1.2 Some Hypotheses Tested

The pattern of the results given in table 6 provides an overview of the effectiveness of the television lectures. The sequences which were either satisfactory or unsatisfactory are quite obvious by following the pattern of the asterisks in column 3 for each lecture. During the television lectures, the percentages were available immediately after the students responded into the Cybercom system. Intervention was never practised even when it was blatantly obvious that the majority of students had not understood the instruction.

The first null hypothesis (see 6.1.3) was tested in terms of the results supplied by table 6. Given that the questions in each lecture were valid, the results in table 6 represent the attainment level achieved by the students of pre-stated behavioural objectives during the presentation of the television lectures, which were supplied via a computer based system. The null hypothesis which claims that this level cannot be assessed using this computer based teaching system was therefore rejected.

The third null hypothesis (section 6.1.3) was reviewed in the l[illegible] of the results available in table 6. It was implied that (section 6.1.2) the lecturer who prepared the lectures considered them to be 'better' than average; it was certainly intended that the vast majority of questions would attain the acceptance level. This was not the case - only twenty three of the fifty seven questions managed to attain this level.

Therefore the null hypothesis which claims there is no significant difference between the actual level achieved by the students and that anticipated by the presenter of the lecture appeared to be false. As this hypothesis was stated as a consequence of the experimental work undertaken and not as a focus of the investigation (see 6.1.3) a statistical test was not performed on this hypothesis.

Nevertheless, this area is worthy of further research as the actual levels of attainment achieved by students during a lecture and the levels which the instructor believes they have attained or will attain appear to be vastly different. Support for this point of view was given by a separate study which was conducted simultaneously with this study and empirical data collected showed similar differences (Young, 1979): Several lecturers were asked to predict the levels of attainment which the students would achieve during the lectures. The predictions were based solely on the considered effectiveness of the instruction and knowledge of the students' capabilities. Statistical tests revealed significant differences between the two levels.

A bar chart is given in figure 14 which shows the distribution of the results given in table 6.

8.2 Analysis and Changes in Instruction

The second stage of the experiment was the analysis of the defective or ineffective instruction. An attempt was made to find what it was that was causing the instruction to be ineffective. Once a possible answer to this problem was found, changes were made to the defective sequences in order to improve their effectiveness.

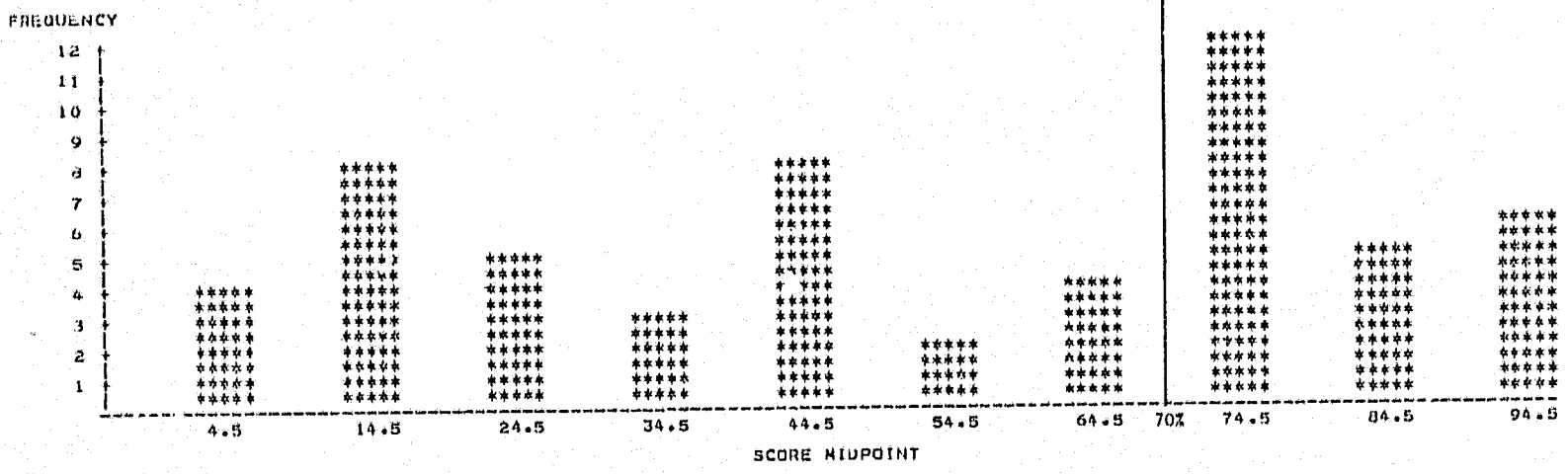
8.2.1 The Basis of the Analysis

When defective instructional sequences are analysed, the implication is made that the instructor was deficient in his instruction during that sequence. He possibly performed in a way in which the 'good' instructor would not have. This introduces the questions relating to the criteria that constitute 'good' instruction, or rather the lack of such accepted criteria as outlined in chapter 6.

FIGURE 14: DISTRIBUTION OF SCORES FOR GROUP 1 (57 QUESTIONS)

(THE DATA WAS OBTAINED FROM TABLE 6)

FREQUENCY BAR CHART



A bibliography of research studies conducted during 1967 on teaching competence revealed that out of 1000 studies, 980 rated teachers in accordance with the impressions or judgements made by supervisors or independent observers (Burkhart, 1969). Frequently, researchers correlated particular teachers' behavioural characteristics with pupil gain measured on achievement tests or self-reports. They analysed teacher behaviour in the classroom in order to find the 'good' teacher. Several criticisms of this method of teacher evaluation have been made. Sorenson and Gross (1967, pp1-2) point out that, 'to attempt to define teaching success in terms of some single fixed teacher ideal is both untenable and inappropriate... teachers are bound to be regarded differently by persons with varying concepts of the teacher's role'.

Only twenty studies in the 1967 bibliography used the criterion of student growth in order to measure teacher effectiveness. It has been noted that two very important reasons why effective and ineffective teachers cannot be described with any assurance are the wide variation in the value concepts underlying the descriptions of desirable teaching objectives and the differences in a teacher's role at different educational levels, in different subjects and with different students (Ryans, 1960). Consistent with this point are the conclusions reported by Solomon, Rosenberg and Bezdek (1964) who state that teachers in the social sciences differ from their counterparts in other areas with respect to certain behavioural dimensions.

Popham (1967, p2), commenting on the classroom observation approach to teacher assessment, concludes that 'the quality of learning which transpires in a given instructional situation is a function of particular instructional procedures employed by a particular instructor for particular students with particular goals in mind'. In the analysis of the defective instruction in this experiment, the objective was to

improve learner growth. The methods employed apply to a particular instructor, sequences and students, although they are derived from methods which enjoy substantial support as will be indicated in the next section.

8.2.2 The Method

When the analysis of the defective instruction was conducted, reasons for students' failure to achieve certain behavioural objectives were sought. A useful method in this regard is the concept of learning hierarchies. This concept has been seen as a useful tool in instructional design, particularly in connection with curriculum sequences and the ordering of learning tasks (Gagné, 1977). A hierarchy of component learning tasks is constructed by beginning with a desired instructional objective and then asking in effect, 'To perform this behaviour, what prerequisite or component behaviours must a learner be able to perform?'. For each behaviour so identified, the same question is asked, thus generating a hierarchy of objectives based on observable prerequisites.

In an earlier work, Gagné and Paradise (1961) suggest that three possible explanations exist when a behavioural objective has not been mastered: a) a subordinate learning set has been omitted, b) a subordinate learning set has not been retained because of inadequate practice, or c) insufficient guidance in the integration of subskills has been provided.

These hierarchies and explanations formed the basis of the method used in the analysis. This is an a posteriori approach to the improvement of the instruction. It could be argued that if this method was employed a priori, then the instruction would have been far more effective initially. The point has been made that not all lecturers are education and teaching specialists and their lectures are usually pre-

pared in a manner in which they subjectively feel will prove effective. This approach was adopted in this study and therefore improvement was attempted a posteriori.

The final point on the method employed in the analysis is who actually conducted the analysis. If a team of experts were employed they would possibly provide a more accurate and precise analysis than if the instructor himself were used. However, the model being investigated is one where the instruction is that of an individual lecturer, and not the effort of a team of experts. The study revolves around the evaluation and improvement of the personal effort of the instructor and whether, using this model the effectiveness can be improved within the given conditions and facilities. Therefore the instructor is employed for this analysis. Support is rendered to this point of view by Bligh (1975, p213): '... I shall argue for the use of self-evaluations. There tends to be an assumption that a teacher is assessed by someone other than himself. I should have thought that assessment of one's own teaching in order to improve it should be both the most important and the most frequent kind of assessment.'

8.2.3 A Summary

Each question which did not achieve the acceptance level (as indicated in table 6) was analysed by reviewing the objective and the instructional sequence which was supposed to have enabled the students to achieve this objective. A detailed description of the entire analysis is given in appendix II, together with similar analyses for groups 2 and 3.

An overview of the analyses of the deficient instructional sequences reveals that the reasons for students not achieving the acceptance level can be placed into one of three categories:

(a) Inadequate instruction. This category refers to instruction which provided insufficient guidance in the integration of subskills and as a result many students were unable to master the behavioural objective, as outlined by Gagné and Paradise (1961). Included in this category is instruction which is based on a dual objective; the instruction might be satisfactory on part of the objective but inadequate on the other (for example see questions 3 and 4, lecture 4).

(b) Subordinate objectives. This category refers to Gagne's (1977) learning hierarchies. The achievement of the behavioural objective is dependent on the mastery of earlier or subordinate objectives. These subordinate objectives sometimes formed the earlier part of the lecture or else were part of previous lectures. In the initial construction of the television lectures it was assumed that the students would have mastered these earlier learning tasks. Included in this category are learning tasks which were not based in previous lectures, but were assumed to have been acquired by the students. An example of this is the assumed student ability of operating an electronic calculator with adequate competence (see questions 5 and 6 in lecture 6).

(c) Unsatisfactory Questions. This category refers to questions which, although they might be valid, were deemed unsatisfactory in that they did not make the intent of the questions clear or they elicited ambiguous answers which were difficult to score (see lecture 1, question 10 and lecture 2, question 5).

Table 7 shows how the questions were grouped into these three categories. An abbreviated notation is used for the question numbers in this table. For example, 4.2 refers to lecture 4, question 2. Only six questions were deemed to be based on inadequate instruction whereas twenty three questions relied on the mastery of subordinate objectives. This appears to imply that the level of the students' mastery of sub-

Table 7: Categorisation of Questions which did not Attain the Acceptance Level

(a) Inadequate Instruction		6
(i) Insufficient guidance: 2.4; 3.5; 5.4; 6.4.	4	
(ii) Dual objective : 4.3; 4.4.	2	
(b) Subordinate Objectives		23
(i) In preceding stages of lecture: 1.6; 1.7; 1.8; 2.6; 2.7; 3.4; 3.7; 3.8; 3.9; 3.10; 6.9.	11	
(ii) In previous lectures : 2.3; 3.2; 4.2; 4.6; 4.7; 4.8; 5.1; 5.2; 6.1; 6.8.	10	
(iii) Assumed ability : 6.5; 6.6.	2	
(c) Unsatisfactory Questions: 1.10; 2.4 (Excluded from experiment)	2	2
		31
Questions Not Categorised (Result close to Acceptance Level) 2.10; 5.10; 6.7.		3
		34

Table 8: Summary of Instructional Changes

(a) Replacement: 2.4.		1
(b) Addition: 1.6; 1.7; 1.8; 2.3; 2.6; 2.7; 3.2; 3.4; 3.5; 3.7; 4.2; 4.3; 4.4; 4.6; 4.7; 4.8; 5.1; 5.2; 5.4; 6.1; 6.4; 6.5; 6.6; 6.8.		24
		25
No Instructional Change		
(i) Result close to acceptance level: 2.10; 5.10; 6.7.		3
(ii) Earlier changes sufficient: 3.8; 3.9; 3.10; 6.9		4
Changed Questions Excluded from Experiment		2
		34

ordinate objectives was well below the level assumed when the lectures were produced . Three questions were not categorised as their results were very close to the acceptance level. Two questions were changed and are therefore excluded from the experiment as results obtained from different questions cannot be compared. This reduces the total number of questions used in the experiment from fifty seven to fifty five.

Changes to the unsatisfactory instructional sequences were made in an attempt to remedy them. The type of change made to the instruction preceding each question is given in table 8. Only one instructional sequence was entirely removed and replaced with a new instructional sequence. Other improvements, to the instruction preceding twenty-four questions, were attempted by adding a further sequence to the existing one. In most cases this represented the review of subordinate learning tasks which the students should already have mastered. This also lengthened the lectures on average by about five minutes.

The sequences preceding seven questions were not changed. Three of these questions yielded results which were very close to the level of acceptance and the other four were based on subordinate objectives which the students should now have mastered if the earlier instructional changes were successful.

Thus the state of affairs at the conclusion of the second stage of the experiment can be described as follows. Fifty five questions were involved in the experiment. Of these, twenty three attained the level of acceptance and thirty two did not. Of these thirty two which did not attain the required level, twenty five were related to instructional sequences which were changed, and seven to sequences that were not. Therefore the attempts to improve the instruction in the six lectures resulted twenty five questions being based on changed instruction of some sort, and thirty questions based on instructional sequences which were left untouched.

This situation is illustrated by the graph in figure 15. A value of 1 is given to the questions which are based on changed instruction and a value of 0 to those based on unchanged instruction. The seven questions which did not attain this level and whose instruction was not changed are easily identified.

The third stage of the experiment examined whether these instructional changes did facilitate the attainment of the objectives by an increased percentage of students.

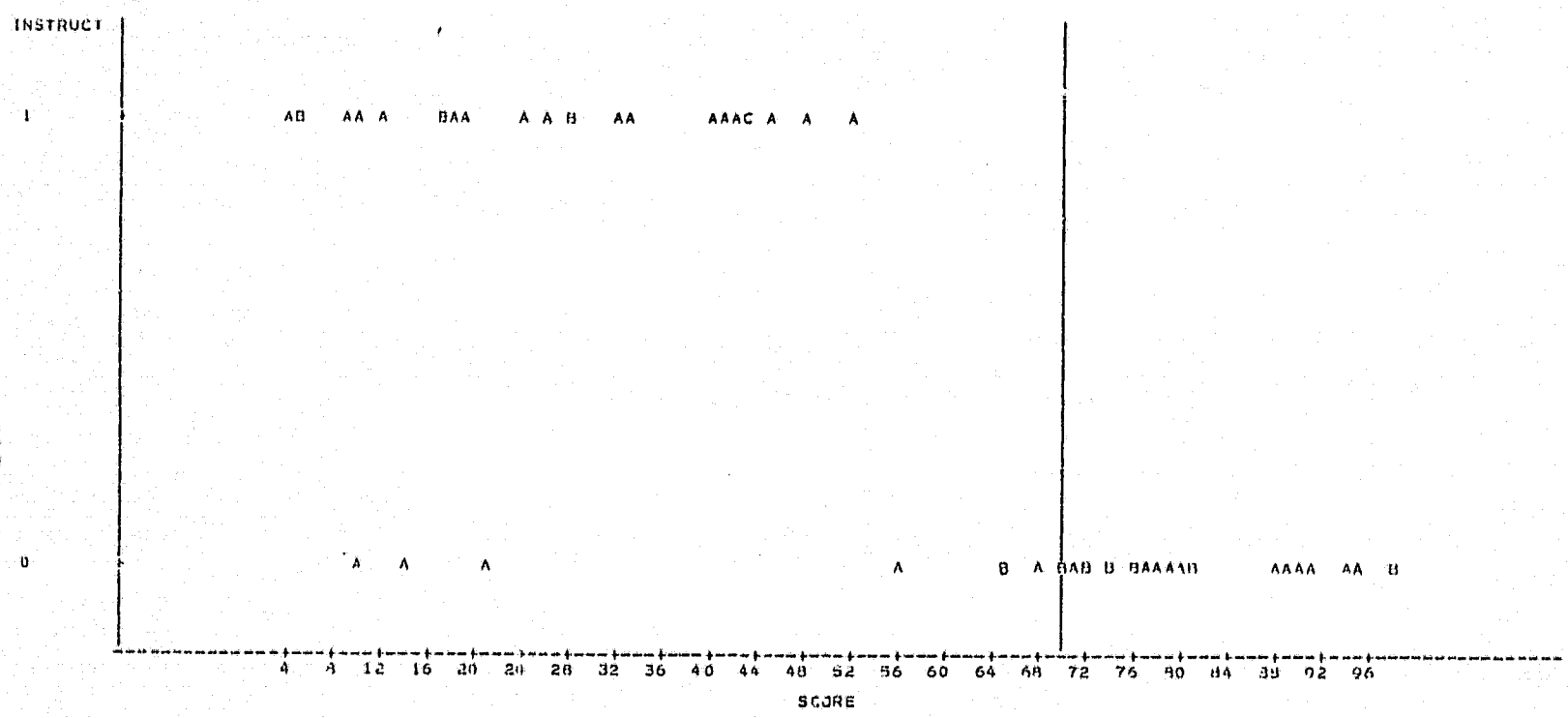
8.3 Final Results

Once the instructional changes had been made, two sets of lectures existed. The first set of six lectures comprised the initial instruction. The second set of six lectures comprised instructional sequences which can be placed into two groups: the first group consisted of instructional sequences which were left untouched and were therefore identical to the initial instruction. Thirty questions were based on these sequences. The second group consisted of instructional sequences which were changed in one way or another and therefore were not identical to the initial instruction. Twenty five questions were based on these sequences. The following terminology will be used in the remainder of the discussion. The instruction upon which the thirty questions were based (the instruction which was identical in both sets of lectures) will be referred to as the unchanged instruction, whereas the sequences which were changed (the instruction differs in the two sets of lectures) will be referred to as the changed instruction. The initial set of six lectures will be referred to as the initial instruction and the second set of six lectures will be referred to as the final instruction.

FIGURE 15: GRAPH OF THE T.V. LECTURE QUESTIONS FOR GROUP 1

KEY : X AXIS - QUESTION SCORES
Y AXIS - 0=UNCHANGED INSTRUCTION
1=CHANGED INSTRUCTION

LEGEND: A = 1 OBS, B = 2 OBS, ETC.



Thus both changed and unchanged instruction can be referred to in the initial instruction, but it must be understood that the changed instruction referred to here proved to be unsatisfactory and was to be changed. The sequences had as yet not undergone any changes and therefore therewere differences between the changed instruction for the initial and final instructional sequences. This arrangement is tabulated in table 9.

Table 9: Relationship between Changed, Unchanged, Initial and Final Instruction

	Initial	Final
Changed	25 Questions	25 Questions
Unchanged	30 Questions	30 Questions
	55 Questions	55 Questions

The third stage of the experiment was concerned with testing the null hypothesis which stated that the level of attainment achieved by students cannot be raised by administering changes to the instruction after the instruction has been analysed. Therefore this section will determine whether or not there were any significant differences in the results obtained for the twenty five questions which were based on instructional sequences which were changed.

8.3.1 The Question Scores

Two more groups were introduced to the experiment at this stage; groups 2 and 3. Group 2 was shown the initial instruction and therefore was subjected to treatment identical to that administered to group 1.

The reason for this was twofold. First, the results of the pretest could be confirmed. The pretest indicated that there were no significant differences between the groups. Confirmation of this can be obtained by comparing the results achieved by groups 1 and 2 during the television instruction. Secondly, this comparison would also act as an indication of the reliability of the questions.

Group 3 were shown the final instruction. This meant that thirty questions in the television lectures were based on instructional sequences which were identical to those seen by both groups 1 and 2; these sequences were unchanged. However, Group 3 were shown instructional sequences upon which twenty five questions were based which were different to the sequences seen by groups 1 and 2.

The results obtained by all three groups for the fifty five questions are given in table 10. Note that a point separates the lecture number from the question number in the question column. This notation will be used in all the result tables in this section.

The results obtained by groups 1 and 2 should not be significantly different as they were subject to identical instruction. Their results are illustrated graphically in figures 16 and 17. The scores obtained by each group are plotted at regular class intervals. The patterns for the two distributions are easily discernable and have a similar appearance.

If the changed instruction had a marked effect on the results obtained by Group 3, then the distribution of these results should not have a similar pattern to the distributions for groups 1 and 2. This is shown to be the case in figure 18.

Although the distributions appear different, this graphical representation is not a stringent indicator that the results obtained by group 3 for the changed instruction were significantly different from groups 1 and 2. The first step to be taken in order to determine this is to con-

TABLE 10: RESULTS OF 55 QUESTIONS FOR GROUPS 1 , 2 AND 3

OBS	QUESTION	GROUP1	GROUP2	GROUP3
1	1.1	72	80	86
2	1.2	91	94	98
3	1.3	74	88	90
4	1.4	94	82	73
5	1.5	90	94	87
6	1.6	9	2	25
7	1.7	28	18	55
8	1.8	33	27	63
9	1.9	89	88	86
10	2.1	98	92	88
11	2.2	81	56	73
12	2.3	18	22	87
13	2.4	26	28	35
14	2.6	43	27	86
15	2.7	12	6	84
16	2.8	78	89	88
17	2.9	77	51	62
18	2.0	68	55	56
19	3.1	70	76	69
20	3.2	5	5	36
21	3.3	74	73	80
22	3.4	24	35	75
23	3.5	43	49	76
24	3.6	80	76	97
25	3.7	19	31	64
26	3.8	50	72	67
27	3.9	14	18	0
28	3.0	10	4	3
29	4.1	79	81	84
30	4.2	40	44	71
31	4.3	41	43	80
32	4.4	48	26	55
33	4.5	72	74	89
34	4.6	5	18	55
35	4.7	4	39	61
36	4.8	32	57	66
37	5.1	45	3	36
38	5.2	43	39	88
39	5.3	76	82	96
40	5.4	52	62	84
41	5.5	76	88	91
42	5.6	98	95	89
43	5.7	70	81	91
44	5.8	71	71	79
45	5.9	88	99	87
46	5.0	65	79	80
47	6.1	42	71	93
48	6.2	81	86	95
49	6.3	95	98	95
50	6.4	10	41	69
51	6.5	17	30	40
52	6.6	28	25	55
53	6.7	65	83	81
54	6.8	17	43	76
55	6.9	21	54	62

FIGURE 16: DISTRIBUTION OF RESULTS FOR GROUP 1 (55 QUESTIONS)

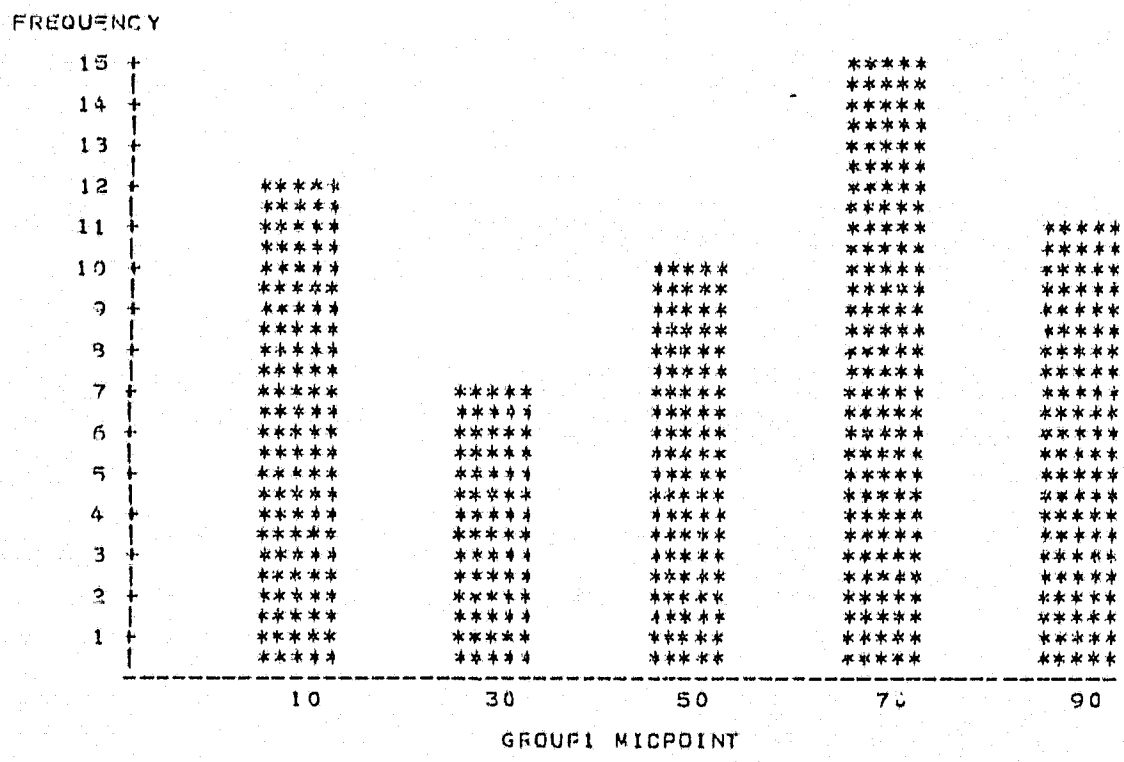


FIGURE 17: DISTRIBUTION OF RESULTS FOR GROUP 2 (55 QUESTIONS)

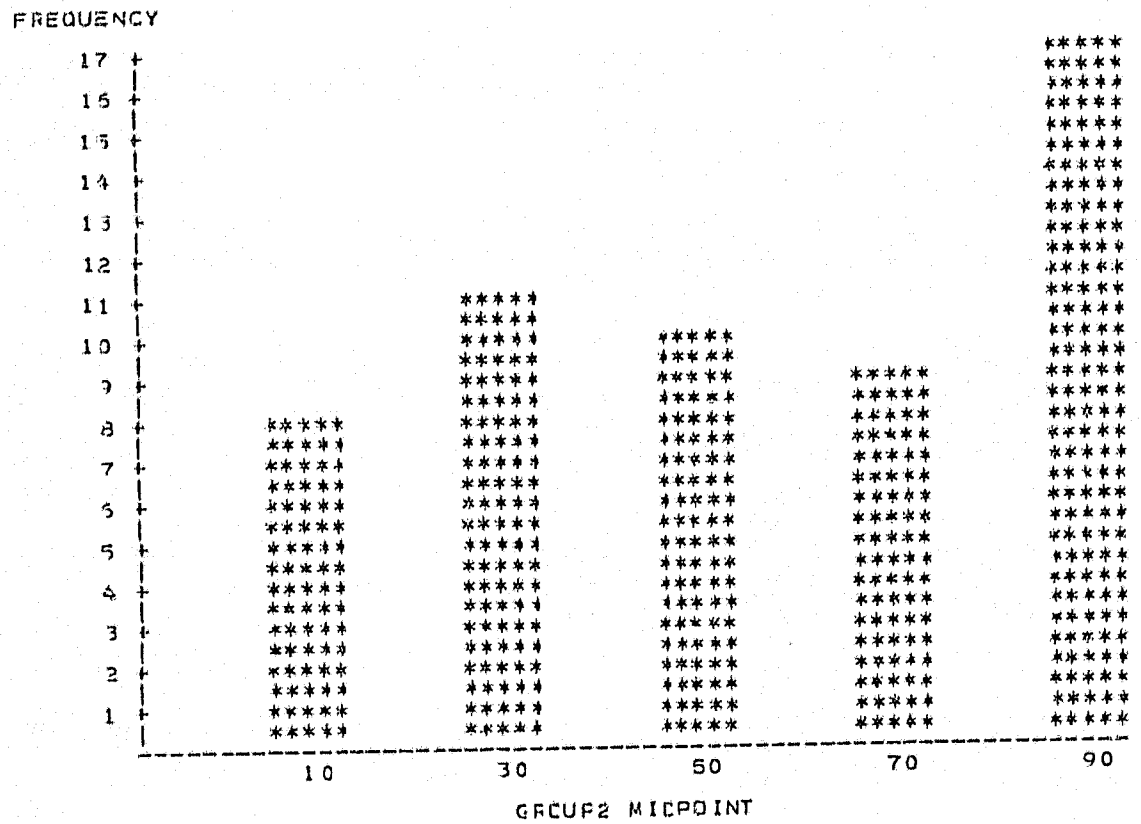
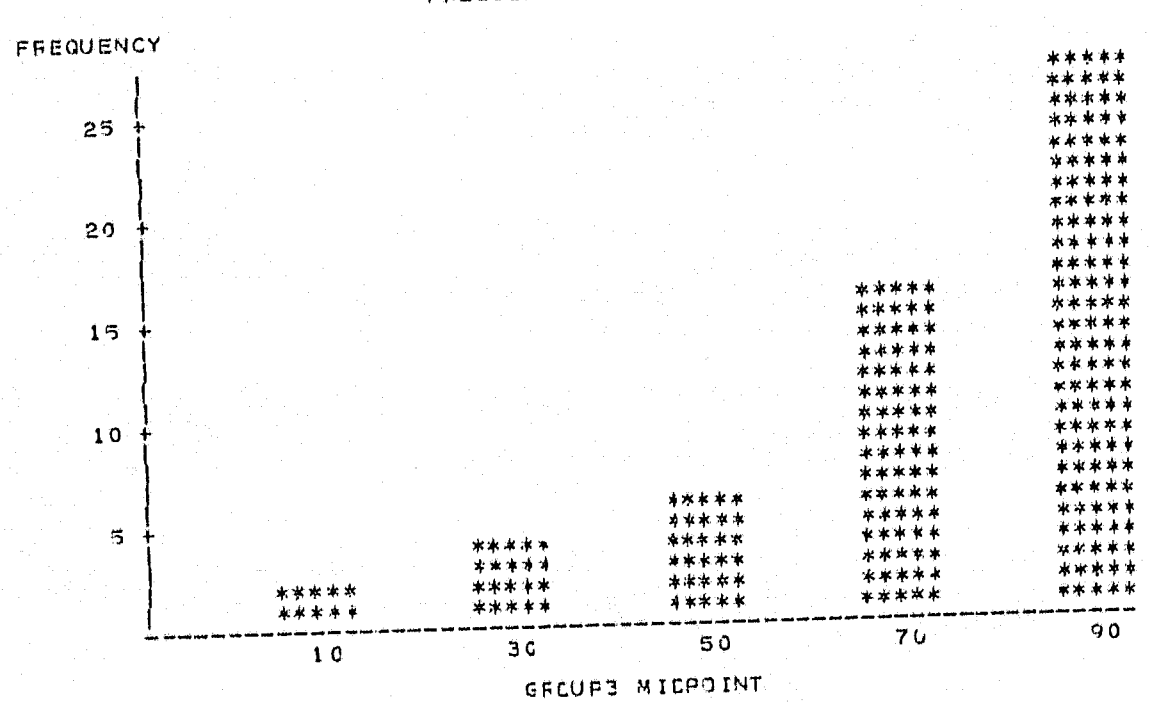


FIGURE 18: DISTRIBUTION OF RESULTS FOR GROUP 3 (55 QUESTIONS)



sider the differences between the scores.

8.3.2 The Differences in Scores

The results which were given in table 10 have been separated into two groups in table 11: those based on unchanged instruction and those based on changed instruction. The notation used in this section to indicate unchanged instruction is a 0, and a 1 for changed instruction. Notice from table 11 that thirty questions are based on unchanged instruction and twenty five on changed instruction.

The scores obtained by the three groups for the unchanged instruction should not be significantly different as the instruction was identical. In other words the differences in the scores should be distributed about zero.

The same argument holds for groups 1 and 2 for the changed instruction. They were both shown the initial lectures, before any sequences were changed. However, if the changes had a influence on the results, then Group 3 should not obtain similar results to groups 1 and 2 for these questions. In other words, the differences in the scores between group 3 and both groups 1 and 2 should not be distributed about zero.

In order to determine this, three sets of differences were found. The first, DIFF1, was obtained by subtracting the score obtained by group 1 from that obtained by group 2 for each of the fifty five questions. The second, DIFF2, by subtracting group 1 scores from group 3 scores and the third DIFF3, by subtracting group 2 from group 3 scores. Therefore:

DIFF1 = GROUP2 - GROUP1

DIFF2 = GROUP3 - GROUP1

DIFF3 = GROUP3 - GROUP2

TABLE 11: RESULTS FOR CHANGED AND UNCHANGED INSTRUCTION

KEY: 0 — UNCHANGED INSTRUCTION
1 — CHANGED INSTRUCTION

INSTRUCT=0

OBS	QUESTION	GROUP1	GROUP2	GROUP3
1	1.1	72	80	86
2	1.2	91	94	98
3	1.3	74	88	90
4	1.4	94	82	73
5	1.5	90	94	87
6	1.9	89	88	86
7	2.1	98	92	88
8	2.2	81	56	73
9	2.8	78	89	88
10	2.9	77	51	62
11	2.0	68	55	56
12	3.1	70	76	69
13	3.3	74	73	80
14	3.6	80	76	97
15	3.8	56	72	67
16	3.9	14	18	0
17	3.0	10	4	3
18	4.1	79	81	84
19	4.5	72	74	9
20	5.3	76	82	.
21	5.5	76	88	91
22	5.6	98	95	89
23	5.7	70	81	91
24	5.8	71	71	79
25	5.9	88	99	87
26	5.0	65	79	80
27	6.2	81	86	95
28	6.3	95	98	95
29	6.7	65	83	81
30	6.9	21	54	62

INSTRUCT=1

OBS	QUESTION	GROUP1	GROUP2	GROUP3
31	1.6	9	2	25
32	1.7	28	18	55
33	1.8	33	27	63
34	2.3	18	22	87
35	2.4	26	28	35
36	2.6	43	27	86
37	2.7	12	6	84
38	3.2	5	5	36
39	3.4	24	35	75
40	3.5	43	49	76
41	3.7	19	31	64
42	4.2	40	44	71
43	4.3	41	43	80
44	4.4	48	26	55
45	4.6	5	18	55
46	4.7	4	39	61
47	4.8	32	57	66
48	5.1	45	3	36
49	5.2	43	39	88
50	5.4	52	62	84
51	6.1	42	71	93
52	6.4	10	41	69
53	6.5	17	30	40
54	6.6	28	25	55
55	6.8	17	43	76

These differences are listed in table 12 and are separated into two sets of results - one for the changed and the other for the unchanged instruction.

An examination of table 12 reveals that all three differences for the unchanged instruction have negative and positive values, indicating that they might well be distributed about zero. However only DIFF1 has positive and negative values for the changed instruction (if the value for DIFF2 at question 5.1 is ignored). DIFF2 and DIFF3 have only positive values. This indicates that group 3 had higher scores than groups 1 and 2 for all the questions based on the changed instruction.

This point is illustrated graphically in figures 19 to 21. Figure 19 shows the distribution of the scores for DIFF1. They are plotted in two sets; those based on the changed and those based on the unchanged instruction. Notice that when the differences are equal only one point is plotted (either a 0 or 1). Therefore thirteen observations are hidden in figure 19. These two sets of instruction were identical for groups 1 and 2. Therefore both sets should be similarly distributed about zero. An examination of figure 19 reveals this to be the case, within a reasonable variation. Group 3 were shown the final instruction where the changes had been implemented. Therefore both DIFF2 and DIFF3 should indicate a distribution about zero for the unchanged instruction (this instruction was identical for all three groups) whereas this should not be the case for the changed instruction, if the results were influenced by the changes. Figures 20 and 21 show this to be the case, again within a reasonable variation. The distributions in both these figures for the changed instruction have shifted in a positive direction. The means for these distributions are given in table 13. Both DIFF2 and DIFF3 have noticeably higher means than DIFF1 for the changed instruction whereas the means for the unchanged instruction are all similar.

TABLE 12: DIFFERENCES FOR CHANGED AND UNCHANGED INSTRUCTION

KEY: 0 - UNCHANGED INSTRUCTION. 1 - CHANGED INSTRUCTION
DIFF1 = GROUP2 - GROUP1
DIFF2 = GROUP3 - GROUP1
DIFF3 = GROUP3 - GROUP2

INSTRUCT=0

OBS	QUESTION	DIFF1	DIFF2	DIFF3
1	1.1	8	14	6
2	1.2	3	7	4
3	1.3	14	16	2
4	1.4	-12	-21	-9
5	1.5	4	-3	-7
6	1.9	-1	-3	-2
7	2.1	-6	-10	-4
8	2.2	-25	-8	17
9	2.8	11	10	-1
10	2.9	-26	-15	11
11	2.0	-13	-12	1
12	3.1	6	-1	-7
13	3.3	-1	6	7
14	3.6	-4	17	21
15	3.8	16	11	-5
16	3.9	4	-14	-18
17	3.0	-6	-7	-1
18	4.1	2	5	3
19	4.5	2	17	15
20	5.3	6	20	14
21	5.5	12	15	3
22	5.6	-3	-9	-6
23	5.7	11	21	10
24	5.8	0	8	8
25	5.9	11	-1	-12
26	5.0	14	15	1
27	6.2	5	14	9
28	6.3	3	0	-3
29	6.7	18	16	-2
30	6.9	33	41	8

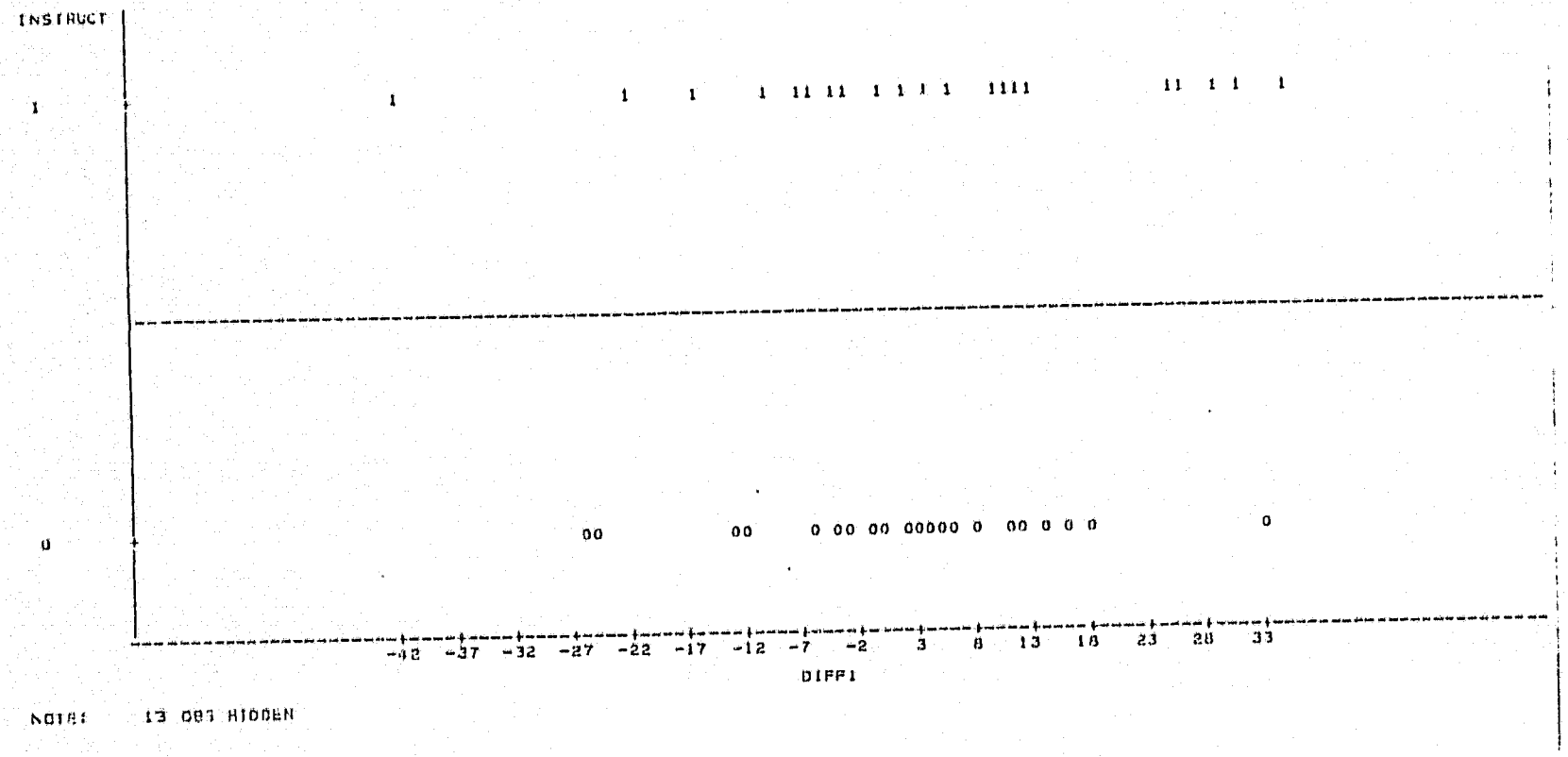
INSTRUCT=1

OBS	QUESTION	DIFF1	DIFF2	DIFF3
31	1.6	-7	16	23
32	1.7	-10	27	37
33	1.8	-6	30	36
34	2.3	4	69	65
35	2.4	2	9	7
36	2.6	-16	43	59
37	2.7	-6	72	78
38	3.2	0	31	31
39	3.4	11	51	40
40	3.5	6	33	27
41	3.7	12	45	33
42	4.2	4	31	27
43	4.3	2	39	37
44	4.4	-22	7	29
45	4.6	13	50	37
46	4.7	35	57	22
47	4.8	25	34	9
48	5.1	-42	-9	33
49	5.2	-4	45	49
50	5.4	10	32	22
51	6.1	29	51	22
52	6.4	31	59	28
53	6.5	13	23	10
54	6.6	-3	27	30
55	6.8	26	59	33

FIGURE 19: DISTRIBUTION OF INSTRUCTION AND DIFF1

-------KEY: X AXIS 0=UNCHANGED INSTRUCTION--------
------- 1=CHANGED INSTRUCTION --------
------- Y AXIS DIFF1=GROUP2 - GROUP1 --------

PLOT OF INSTRUCT*DIFF1 SYMBOL IS VALUE OF INSTRUCT

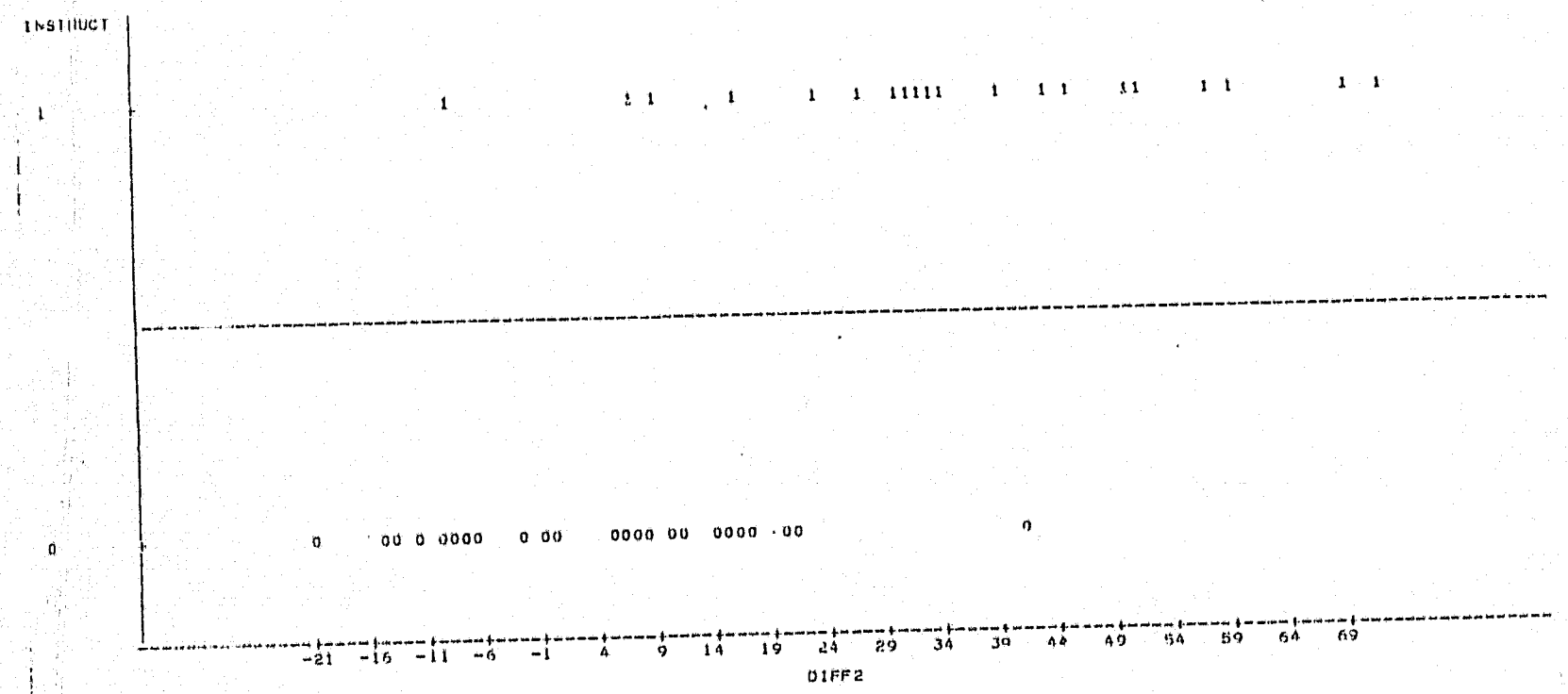


NOTE: 13 OBS HIDDEN

FIGURE 20: DISTRIBUTION OF INSTRUCTION AND DIFF2

------KEY: X AXIS 0=UNCHANGED INSTRUCTION-------
------ 1=CHANGED INSTRUCTION --------
------ Y AXIS DIFF2=GROUPJ - GROUPI --------

PLOT OF INSTRUCT*DIFF2 SYMBOL IS VALUE OF INSTRUCT

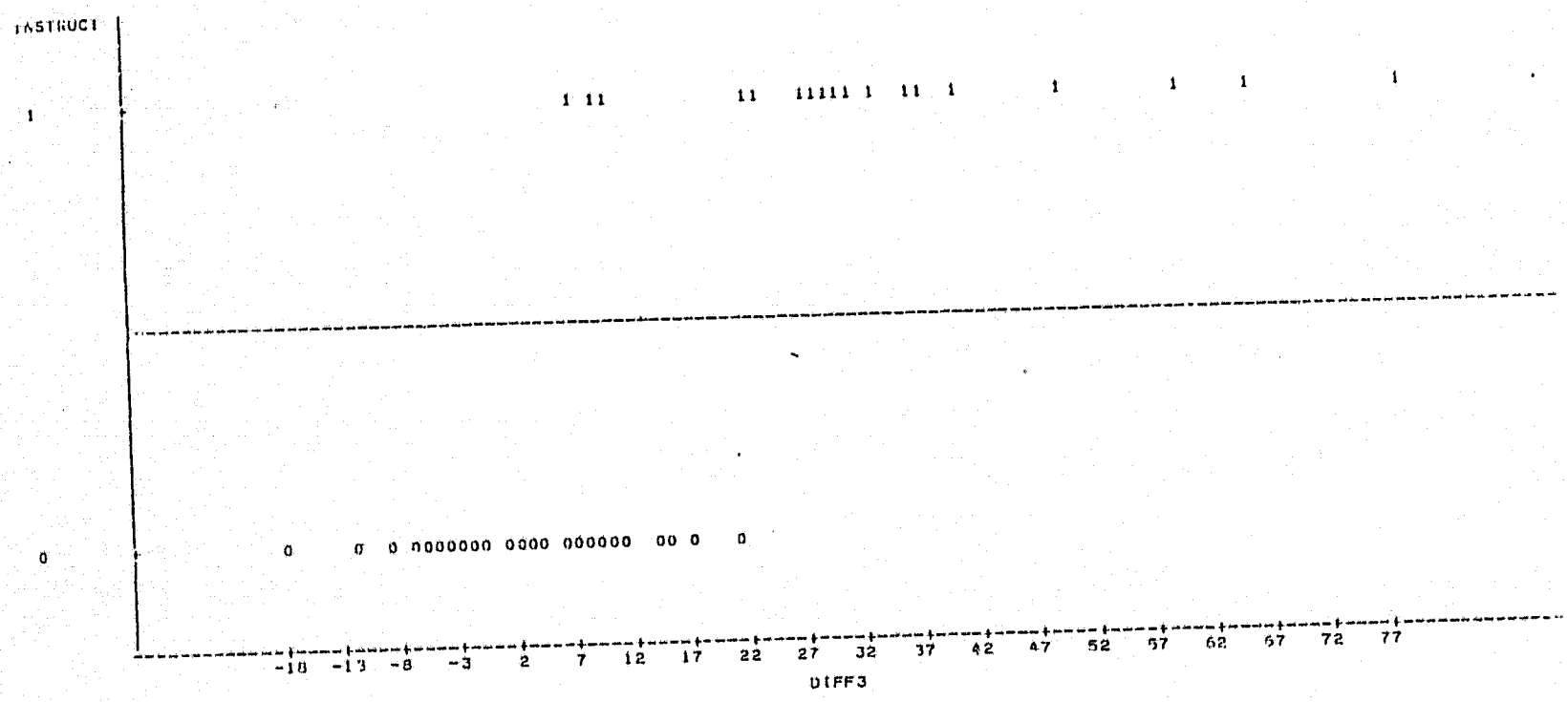


NOTE: 11 OBS HIDDEN

FIGURE 21: DISTRIBUTION OF INSTRUCTION AND DIFF3

KEY: X AXIS 0=UNCHANGED INSTRUCTION
1=CHANGED INSTRUCTION
Y AXIS DIFF3=GROUP3 - GROUP2

PLOT OF INSTRUCT*DIFF3 SYMBOL IS VALUE OF INSTRUCT



NOTE: 15 OBS HIDDEN

Table 13: Means for DIFF1, DIFF2 and DIFF3

	Instruction	
	Unchanged (30 obs)	Changed (25 obs)
DIFF1	2,87	4,28
DIFF2	4,97	37,24
DIFF3	2,10	32,96

The indication is that the scores obtained by Group 3 for the changed instruction were substantially higher than those obtained by groups 1 and 2. Therefore the instructional changes that were implemented must be responsible for this increase: the instruction was the only variable that was permitted to change during the experiment. This result would tend to reject the null hypothesis that stated that an increased number of students would not be able to attain the pre-stated objectives, if the instruction was analysed and changed accordingly.

Before this hypothesis can be rejected, a measure of the significance of these changes must be determined. Two significance tests are conducted in the next two sub-sections in order to determine this. No assumptions were made concerning the distributions of the populations of which DIFF1, DIFF2 and DIFF3 are three samples. Therefore the significance tests that are used are both non-parametric tests.

8.3.3 The Tukey Test

In this and the following section tests are conducted to determine any significant differences between the scores obtained for questions based on the changed and unchanged instruction for the three samples, DIFF1, DIFF2 and DIFF3.

The first test to be administered is the Tukey Test. John Tukey's quick test of location has been described as one of the simplest and quickest significance tests for comparing independent samples (Kogan, 1960). Basler and Smawley (1968), in a study which compares Tukey's Compact Test with various other classic tests support this view and claim that this test has not received the attention it deserves. Their study concludes that the comparisons obtained a very high frequency of agreement with classic techniques. The compactness of the technique however does somewhat reduce its power. As the test does not appear to be as widely used as some classic tests a brief dessription of the test will precede its application to DIFF1, DIFF2 and DIFF3.

Given a pair of samples (where sample 1 contains n_1 observations and sample 2 contains n_2 observations), if n_1 contains the highest value and n_2 the lowest value, then one adds the number of values in n_1 which exceed all those values in n_2, with the number of values in n_2 which fall below all those measures in n_1 (Tukey, 1959). The sum is then compared with Tukey's critical values:

If the two groups are of approximately the same size (the ratio of the sizes must not exceed 4:3), then Tukey's critical values for the obtained sum are:

A. A critical value of 7 for 2-tail 5% level (0,05)

B. A critical value of 10 for 2-tail 1% level (0,01)

C. A critical value of 13 for 2-tail 0,1% level (0,001)

A Tukey formula does exist for adjustments to be made to these critical values if the ratio of the sizes exceeds 4:3. This is not necessary for this analysis as the number of questions being compared in the two samples gives a ratio of 30:25 which is less than 4:3.

Tukey Test for DIFF1. The pre-test analysis revealed that there was no significant difference between groups 1 and 2 at the 0,05 significance level. Therefore the differences in their scores for the questions in the television instruction should also reveal no significant difference between the changed and unchanged instruction, as all the instructional sequences were the original, untouched versions.

The distribution of DIFF1 for the changed and unchanged instruction is given in figure 22. Note that the notation is used where A represents one observation, B two, C three and so on. If figure 22 is closely examined it will be seen that the changed instruction (value of 1) contains both the highest and lowest values of 35 and -42 respectively. The whole range of values for the unchanged instruction lies between these two values.

Therefore one cannot even suggest that there is a significant difference because an initial requirement for significance is that the sample containing the highest score does not also contain the lowest. This result is therefore as expected: no significant difference.

Tukey Test for DIFF2: The distribution of DIFF2 for the changed and unchanged instruction is given in figure 23. Note from this figure that the changed instruction has eleven values greater than the unchanged instruction, and the unchanged instruction has five values smaller than the changed instruction. Thus the statistic is:

$$11 + 5 = 16.$$

At the 0,001 level, the Tukey critical value is 13. Therefore the results for the changed instruction were significantly different. This means that group 3 scored significantly higher on the changed instruction at the 0,001 level.

FIGURE 22: GRAPH FOR TUKEY TEST ON DIFF1

-------KEY: X AXIS 0=UNCHANGED INSTRUCTION--------
-------   1=CHANGED INSTRUCTION --------
-------   Y AXIS DIFF1=GROUP2 - GROUP1 --------

PLOT OF INSTRUCT*DIFF1 LEGEND: A = 1 OBS, B = 2 OBS, ETC.

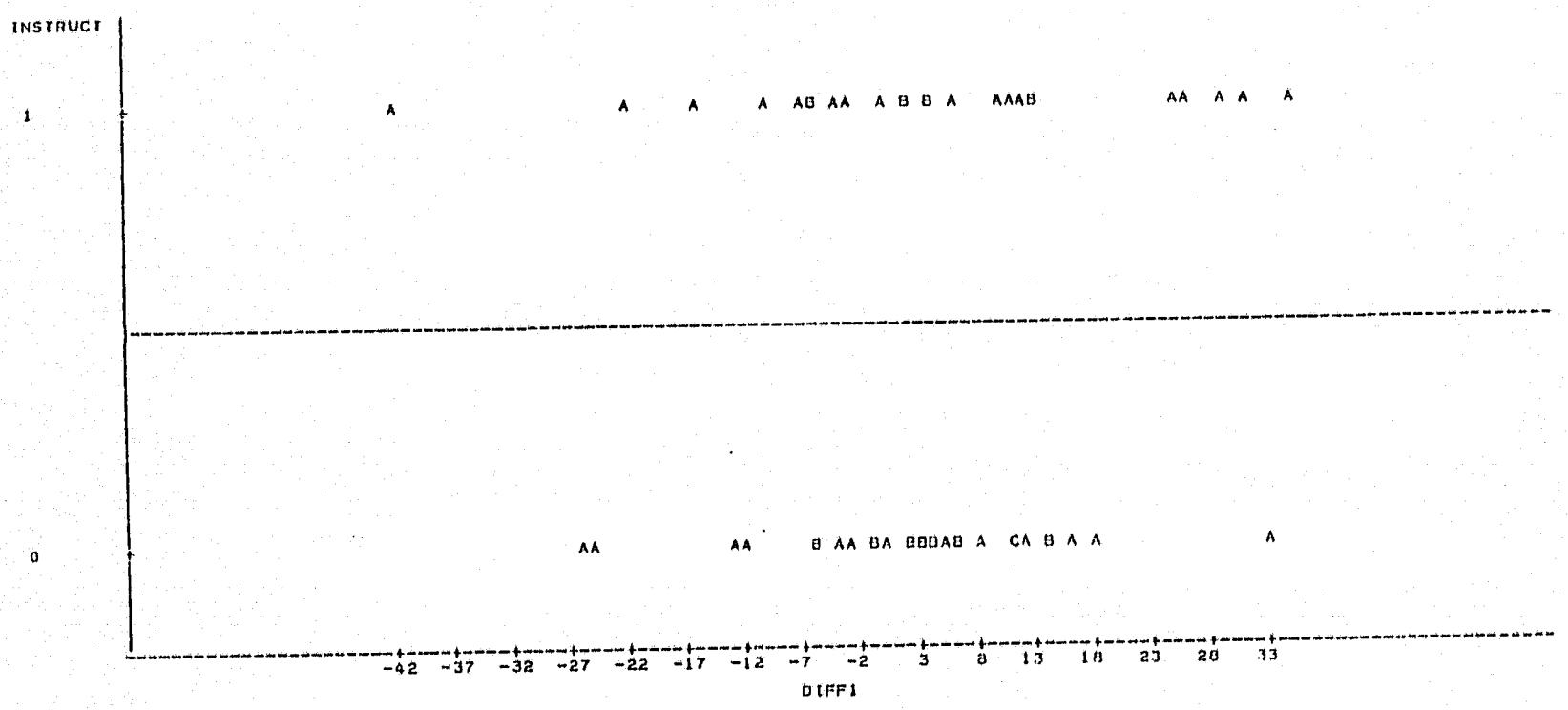
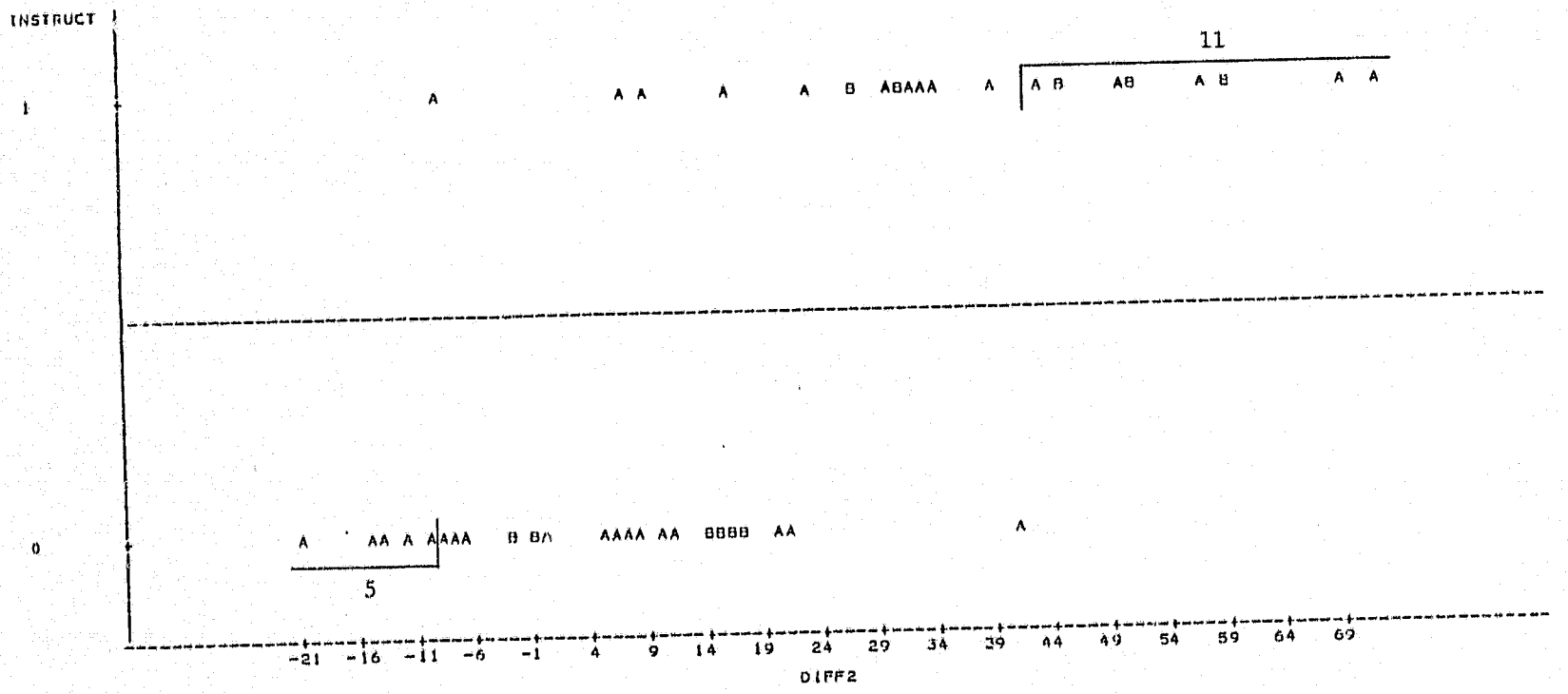


FIGURE 23: GRAPH FOR TUKEY TEST ON DIFF2

------KEY: X AXIS 0=UNCHANGED INSTRUCTION--------
------ 1=CHANGED INSTRUCTION --------
------ Y AXIS DIFF2=GROUP3 - GROUP1 --------

PLOT OF INSTRUCT*DIFF2 LEGEND: A = 1 OBS, B = 2 OBS, ETC.



Tukey Test for DIFF 3: The result for DIFF3 should be similar to that obtained for DIFF2 as outlined earlier. Therefore the critical value thirteen for the Tukey test should be exceeded. Figure 24 gives the distribution of the changed and unchanged instruction for DIFF3. Examination of the figure reveals a value for the statistic of forty-two which is certainly significant at the 0,001 level.

8.3.4 The Mann-Whitney U Test

The Mann-Whitney test is a nonparametric test which uses the sum of ranks of scores in hypothesis testing (Chao, 1974). Therefore, no assumptions are made about the populations from which the samples are taken, as is the case with the three samples in this analysis, DIFF1, DIFF2 and DIFF3. This test is also referred to as the U test, since the test statistic U is computed from the sample data. The method is as follows.

Say we are considering that two independent samples, 1 and 2 (with n_1 and n_2 observations respectively) are to be tested for differences from some population mean. A rank is assigned to each score according to its magnitude. Then R_1 is the sum of the ranks for sample 1 and R_2 the sum for sample 2. The statistic U is then defined as:

$$U = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1 \text{ , or alternately,}$$

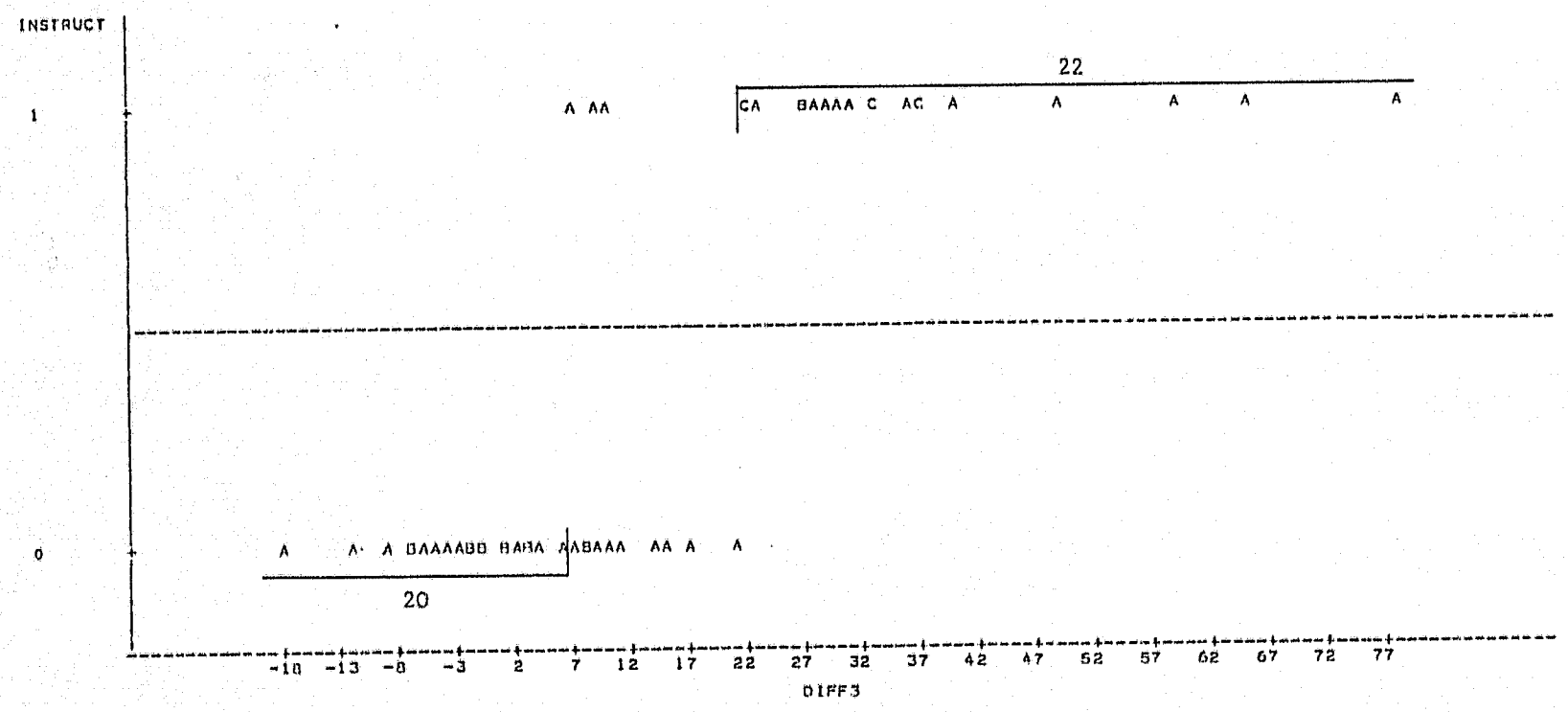
$$U' = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2 \text{ .}$$

The two U values are related by $U = n_1 n_2 - U'$. The smaller U value should be used for greater confidence.

FIGURE 24: GRAPH FOR TUKEY TEST ON DIFF3

------KEY: X AXIS 0=UNCHANGED INSTRUCTION-------
------ 1=CHANGED INSTRUCTION -------
------ Y AXIS DIFF3=GROUP3 - GROUP2 -------

PLOT OF INSTRUCT*DIFF3 LEGEND: A = 1 OBS, B = 2 OBS, ETC.



If n_1 or n_2 are greater than twenty, U is approximately normal with a mean,

$$E(U) = \frac{n_1 \cdot n_2}{2}$$

and a standard deviation,

$$\sigma_u = \sqrt{\frac{n_1 n_2 (n_1+n_2+1)}{12}} \quad .$$

In this analysis, $n_1 = 30$ and $n_2 = 25$. Thus U will be almost normally distributed with a mean,

$$E(U) = \frac{n_1 \cdot n_2}{2} = \frac{(30)(25)}{2} = 375$$

and a standard deviation,

$$\sigma_u = \sqrt{\frac{n_1 n_2 (n_1+n_2+1)}{12}} = \sqrt{\frac{(30)(25)(56)}{12}}$$

$$= 59,16.$$

The Mann-Whitney U test will now be applied to the scores for DIFF1, DIFF2 and DIFF3, in order to determine any significant differences between the changed and unchanged instruction.

U Test for DIFF1: The rankings for DIFF1 are given in table 14.

From this table, $R_1 = 829,5$

$$\therefore \quad U = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1$$

$$= (30)(25) + \frac{30(31)}{2} - 829,5$$

$$= 385,5$$

$$U' = n_1 \cdot n_2 - U = 364,5 \quad .$$

$$\therefore \quad \text{Use } U = 364,5.$$

Table 14: Ranking of DIFF1

Instruct = 0			Instruct = 1		
Question	Diff1	Rank	Question	Diff1	Rank
2.9	-26	2	5.1	-42	1
2.2	-25	3	4.4	-22	4
2.0	-13	6	2.6	-16	5
1.4	-12	7	1.7	-10	8
2.1	-6	11,5	1.6	-7	9
3.0	-6	11,5	1.8	-6	11,5
3.6	-4	14,5	2.7	-6	11,5
5.6	-3	16,5	5.2	-4	14,5
1.9	-1	18,5	6.6	-3	16,5
3.3	-1	18,5	3.2	0	20,5
5.8	0	20,5	2.4	2	23,5
4.1	2	23,5	4.3	2	23,5
4.5	2	23,5	2.3	4	29,5
1.2	3	26,5	4.2	4	29,5
6.3	3	26,5	3.5	6	34
1.5	4	29,5	5.4	10	37
3.9	4	29,5	3.4	11	39,5
6.2	5	32	3.7	12	42,5
3.1	6	34	4.6	13	44,5
5.3	6	34	6.5	13	44,5
1.1	8	36	4.8	25	50
2.8	11	39,5	6.8	26	51
5.7	11	39,5	6.1	29	52
5.9	11	39,5	6.4	31	53
5 5	12	42,5	4.7	35	55
1.3	14	46,5			R_2=710,5
5.0	14	46,5			
3.8	16	48			
6.7	18	49			
6.9	33	54			
		R_1=829,5			

The z score is calculated from the equation,

$$z = \frac{U - E(U)}{\sigma_u}$$

$$= \frac{(364,5 - 375)}{59,16}$$

$$= -0,177.$$

There is no significant difference for a two tailed test if $-3,27 \leq z \leq 3,27$, at the 0,001 significance level (Chao, 1974). Thus there is no significant difference between the changed and unchanged instruction for DIFF1, which agrees with the Tukey test result.

U Test for DIFF2: The rankings for DIFF2 are given in table 15. From this table, $R_1 = 533$.

$$\therefore \quad U = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1$$

$$= (30)(25) + \frac{30(31)}{2} - 533$$

$$= 682$$

$$U' = n_1 . n_2 - U = +68$$

$\therefore$ Use $U = +68$.

The z score is calculated from the equation,

$$z = \frac{U - E(U)}{\sigma_u}$$

$$= \frac{(+68 - 375)}{59,16}$$

$$= -5,189.$$

Table 15: Ranking of DIFF2

Instruct = 0			Instruct = 1		
Question	Diff2	Rank	Question	Diff2	Rank
1.4	-21	1	5.1	-9	6,5
2.9	-15	2	4.4	7	17,5
3.9	-14	3	2.4	9	20
2.0	-12	4	1.6	16	28
2.1	-10	5	6.5	23	34
5.6	-9	6,5	1.7	27	35,5
2.2	-8	8	6.6	27	35,5
3.0	-7	9	1.8	30	37
1.5	-3	10,5	3.2	31	38,5
1.9	-3	10,5	4.2	31	38,5
3.1	-1	12,5	5.4	32	40
5.9	-1	12,5	3.5	33	41
6.3	0	14	4.8	34	42
4.1	5	15	4.3	39	43
3.3	6	16	2.6	43	45
1.2	7	17,5	3.7	45	46,5
5.8	8	19	5.2	45	46,5
2.8	10	21	4.6	50	48
3.8	11	22	3.4	51	49,5
1.1	14	23,5	6.1	51	49,5
6.2	14	23,5	4.7	57	51
5.5	15	25,5	6.4	59	52,5
5.0	15	25,5	6.8	59	52,5
1.3	16	28	2.3	9	54
6.7	16	28	2.7	72	55
3.6	17	30,5			R_2=1007,0
4.5	17	30,5			
5.3	20	32			
5.7	21	33			
6.9	41	44			
		R_1=533,0			

As there is no significant difference at the 0,001 level for a two-tailed test if $-3,27 \leq z \leq 3,27$, there is a significant difference at this level between the changed and unchanged instruction for DIFF2, which is again consistant with the Tukey test result.

U Test for DIFF3: The rankings for DIFF3 are given in table 16. From this table, $R_1 = 486,5$.

$$\therefore \quad U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1$$

$$= (30)(25) + \frac{30(31)}{2} - 486,5$$

$$= 728,5$$

$$U' = n_1 . n_2 - U = 21,5$$

$$\therefore \quad \text{Use } U = 21,5.$$

The z score is calculated from the equation,

$$z = \frac{U - E(U)}{\sigma_u}$$

$$= \frac{21,5 - 355}{59,16}$$

$$= -5,637.$$

This result is similar to that obtained for DIFF2. Therefore there is a significant different at the 0,001 level between the changed and unchanged instruction for DIFF3.

Table 16: Ranking of DIFF3

Instruct = 0			Instruct = 1		
Question	Diff3	Rank	Question	Diff3	Rank
3.9	-18	1	2.4	7	21,5
5.9	-12	2	4.8	9	25,5
1.4	-9	3	6.5	10	27,5
1.5	-7	4,5	4.7	22	35
3.1	-7	4,5	5.4	22	35
5.6	-6	6	6.1	22	35
3.8	-5	7	1.6	23	37
2.1	-4	8	3.5	27	38,5
6.3	-3	9	4.2	27	38,5
1.9	-2	10,5	6.4	28	40
6.7	-2	10,5	4.4	29	41
2.8	-1	12,5	6.6	30	42
3.0	-1	12,5	3.2	31	43
2.0	1	14,5	3.7	33	45
5.0	1	14,5	5.1	33	45
1.3	2	16	6.8	33	45
4.1	3	17,5	1.8	36	47
5.5	3	17,5	1.7	37	49
1.2	4	19	4.3	37	49
1.1	6	20	4.6	37	49
3.3	7	21,5	3.4	40	51
5.8	8	23,5	5.2	49	52
6.9	8	23,5	2.6	59	53
6.2	9	25,5	2.3	65	54
5.7	10	27,5	2.7	78	55
2.9	11	29			R_2=1053,5
5.3	14	30			
4.5	15	31			
2.2	17	32			
3.6	21	33			
		R_1=486,5			

The significance tests performed in this section show that group 3 scored significantly better, at the 0,001 level, than groups 1 and 2 for the changed television instruction. As group 3 was the only group to view the final instruction where changes had actually been implemented, the conclusion can be drawn that these changes made the instruction significantly better in enabling the students to achieve the pre-stated objectives.

The null hypothesis which stated that, the level of attainment achieved by students of pre-stated behavioural objectives during the presentation of recorded television lectures on introductory statistics could not be raised by analysing the effectiveness of the television lectures and administering changes to parts of the lectures, can be rejected with 99,9% confidence.

9. CONCLUSION

This study has involved itself in three distinct areas which form part of the systems approach to education. The first of these is the area of instructional intent. The ends of the television lectures were clearly stated in terms of observable behavioural objectives. The second area is that of evaluation. The effectiveness of the lectures was measured in terms of their stated intentions. The third area within which the study has moved is that of instructional improvement. An analysis of the methods, objectives and behavioural state of the students took place with a view to improving the instruction. Some concluding remarks will be made in each of these three areas.

9.1 The Instructional Intent

It was stated at the outset of this study that the focus was somewhat limited; only one aspect of the educational programme designed for the introductory statistics course was evaluated. The approach adopted in this aspect, was one which asked how effective lectures were in achieving their intended purpose, given that lectures form an integral part of the programme.

In efforts to ensure instructional effectiveness, producers have done much to obtain professional judgement of instructional television programmes, but little to determine the value of such judgement (Lucas, 1974). It would appear from this study that this judgement is often badly placed. Learner validation of the instruction does not always agree with teacher evaluation. However in both cases, evaluation cannot

be measured in absolute terms unless it is clear what it is that the instruction hopes to achieve.

The whole basis of this study revolved around the stated behavioural objectives of the instruction and this indicates the importance of pre-stated objectives, if any sort of meaningful evaluation of the instruction is to take place. This approach was introduced to the lectures with relative ease as the subject lends itself to the cognitive rather than the affective areas. Thus although the importance of specific behavioural objectives for every sequence of the instruction is stressed, this conclusion could not be applied to other subjects which operate more in the affective area, without further research being pursued in this direction.

One shortcoming of this work was that it did not attempt to analyse the type or level of the stated objectives. If the research was to be repeated and it was found that all the objectives were aimed at the lowest cognitive level, some should be changed to ensure that objectives of a better 'type' were present.

What this research did reveal was that all too often students had not achieved earlier behavioural objectives and this was the source of the difficulty in attaining a present objective. A criticism of this work could be made on this point in that this fact could have been revealed by conducting an appropriate pre-test before each lecture. Although this criticism is accepted it must be remembered that the experimentation tried to assimilate the real situation as closely as possible, which is one where assumptions are made as to the student's entry behaviour at the start of a lecture.

However this criticism could be taken a step further: once it was known that in many instances the students' had not achieved previous objectives, was it correct to 'improve' the lecture by inserting reviews of these objectives? An alternative could have been to state the

required entry behaviour for each lecture and to ensure that students were given sufficient appropriate practice in order to master these behaviours between one lecture and the next. This would necessitate involving the other components of the course such as the tutorials, which was not possible in this study as the effectiveness of the television lectures was examined in a situation divorced from the other components.

Further studies in this area could therefore include some integration of the other teaching components in the course. If these components were employed in such a way so as to ensure that students had achieved all the entry behaviours before the lecture began, a more revealing study could be made of the effectiveness of the instruction in achieving the stated objectives for each lecture.

9.2 The Evaluation

Two central features in the evaluation of the effectiveness of the lectures were the insertion of validated questions into the lectures and the immediate recording of the student responses. This ensured that (i) the questions were measuring what they were supposed to measure, and therefore an accurate indication of the level of attainment of the behavioural objectives was obtained and (ii) no other variables interacted with the instruction before the responses were recorded. This method is considered unique and the results, in addition to being very easy to monitor and collect, are a true measure of the immediate success the lecture has had in achieving its objectives. If these validated questions were not inserted into the lectures, this measure would not have been obtained; the question arises as to whether these lectures could have been accurately evaluated at all.

In addition to the results supplied by this method, it has provided a means of self-evaluation for the instructor, in terms of what the students have learnt from the instruction. Facilities are widespread in higher education for lecturers to be able to watch their lectures on video tape either privately or in company with colleagues and students (University Teaching Methods Unit, 1978). How are lecturers going to judge their performances if use is made of these facilities? If it is in terms of learner growth then some type of question or evaluative means must be inserted at different stages of the lecture so as to measure student growth.

This points to a further criticism which could be levelled at this study. If learner growth is the criterion for effective instruction, then (although it is important to attempt measurement of this growth during the instruction) this growth should be measured at certain fixed intervals after the instruction has taken place in order to determine the degree of retention of these behaviours. Research should be initiated in order to determine any relationships between instruction which is considered to be 'immediately' effective and instruction which is 'retentionally' effective. In order to measure the latter, ways of controlling other learning variables during the delay between instruction and measurement would have to be found.

9.3 The Improvement

The results obtained from the experiment have indicated that student achievement of behavioural objectives during a lecture can be improved. Thus although the instructional means used were not always successful (as indicated in appendix II) a development in the effectiveness of the television instruction has taken place. Cronback and Suppes (1969) explain that the term development refers to the production of materials or pro-

cedures based on general ideas of what will be successful.

Further research is required if further development is to take place. The means of instruction should be re-examined in order to determine why some sequences did not enable the students to achieve the required level of attainment after they had been changed and the explanations more fully clarified. An examination is necessary into the effect of raising the level of acceptance with a view to determining if some optimal level exists; and if so what this optimal level is. Although a third world country, South Africa possesses the technical capability to produce television instruction of a high calibre on a large scale. A large majority of the population suffer from a severe teacher shortage and are desperately trying to improve the quality of their education. If television instruction was to be used as part of an attempt to alleviate this situation, it would seem that more work should be initiated in the area of instructional effectiveness and the optimal levels which can be achieved and which should be aimed for in television teaching.

The teaching model developed here has not made dramatic changes from the traditional lecture. Therefore it should be a relatively easy transition for a lecturer to make if he was interested in determining the actual effectiveness of his teaching. Academic staff are not interested in the professional literature that deals with educational research and innovation; they have neither the time nor the patience to read journals written in the jargon of the educationist, but they are interested in that condensed body of knowledge that has immediate applicability to their own problems (Meyer, 1979). Although this instructional model is a relatively compact means of evaluating instruction, insufficient information is supplied as to why an instructional sequence is ineffective. This area could be further researched with the aid of the computer based teaching system which

can supply an almost unlimited amount of feedback concerning student responses.

More knowledge of how we learn from television would result from the development and analysis of why some instructional sequences prove to be ineffective. Sceiford (1978) stresses the importance of this type of research in television for the future.

'Otherwise, we will not, in the next decade, progress with the sophistication necessary to go beyond the hundreds of past studies concerned with the questions "Does television teach?" and "Does it teach better than x?" (Sceiford, 1978, p56).

Appendix I

The Questions and Objectives of the Television Lectures

The questions and objectives for each of the six television lectures (M_{35}-M_{40}) are given in this appendix. The Questions are arranged sequentially in the order in which they occur in the televisir ectures. The objectives are not arranged in the order in which they are to be achieved during each lecture. This is a copy of the list of questions and objectives which was used by the content specialists to validate the questions. A copy of the response sheet used by the content specialists precedes the questions and objectives.

Matching questions to objectives for Lectures M_{35}-M_{40}

Read through the objectives and questions supplied for each lecture. Then put down next to each objective (in the space provided below) the number(s) of the question(s) which you consider to relate to that objective.

More than one question can be related to one particular objective and some questions might not be related to any objective. The objectives and questions are arranged in a random order.

Lecture	Objective	Questions relating to this objective
M_{35}	A	
	B	
	C	
M_{36}	A	
	B	
	C	
	D	
	E	
M_{37}	A	
	B	
	C	
	D	
M_{38}	A	
	B	
	C	

Lecturer	Objective	Questions relating to this objective
M_{39}	A	
	B	
	C	
	D	
M_{40}	A	
	B	
	C	
	D	
	E	
	F	

M_{35}: Probability Distributions Lecture 1

Objectives:

A. Students must be able to distinguish between a discrete probability distribution (probability function) and a continuous probability istribution (probability density function).

B. Students must be able to identify and distinguish between the concepts: random events, random variable, sample space, probability tree and probability distribution.

C. Students must be able to calculate the probability of a particular random event occuring in a sample space and be able to interpret notations such as P(X=3) and f(2) in these calculations.

Questions:

1. How many random events are there in a situation or experiment where a coin is spun once?
2. How many random events are there in a situation or experiment where a die is rolled once?
3. How many random events are there in a situation or experiment where a card is drawn from a pack of well shuffled cards?
4. Consider on experiment where we spin a coin four times. We define a random variable X to be equal to the number of heads we obtain when spinning a coin four times. What is the maximum value that X can have?
5. Consider the same experiment as the one described in question 4. What is the minimum value that X can have?
6. Consider a situation where a coin is spun three times. The random variable X is defined to be equal to the number of heads which occur. What is P(X=2) equal to?

7. Consider a situation where a coin is spun four times. The random variable X is defined to be equal to the number of heads which occur. What is P(X=4) equal to?

8. Consider the same situation as the one described in question 6. What is f(3) equal to?

9. Consider a situation where we refer to a probability function. Will the random variable involved be discrete or will it be continuous?

10. If we are considering a probability density function, would this be a feasible statement to make: P(X=2) = 0,2?

M_{36}: Expected value of a Random Variable Lecture 2

Objectives:

A. Students must be able to identify the symbols used to denote the mean and standard deviation of a sample and of a population.

B. Students must be able to calculate the mean of a sample and the mean of a population.

C. Students must know that the mean of a population and the expected value of a random variable are equivalent.

D. Students must be able to distinguish between the relative frequency of occurence $(\frac{f}{n})$ of an event (where the number of events n is finite) and the probability of occurence $(\frac{f}{n})$ of an event (where the number of events n is infinite.

E. Students must be able to evaluate the expected value of (i) a random variable and (ii) functions of a random variable.

Questions:

1. The mean of a population is represented by the greek letter μ. What is the name of this greek letter?

2. One of the parameters of a population is described by the greek letter σ. What parameter of a population does σ represent?

3. If you are given a set of data as follows:

x_i	f_i
10	3
4	3
6	2
9	1

, what is $\bar{x}$ in this case?

4. Consider the expression $\lim_{n\to\infty} \frac{f}{n}$. What quantity or what concept does this expression define?

5. The mean of a population $\mu = E(X)$. How do we read the expression $E(X)$?

6. Consider a situation where a coin is spun twice. X, the random variable, is equal to the number of heads which occur. What is $E(X)$ equal to?

7. Consider the same situation as the one described in question 6. What is $E(X^2)$ equal to?

8. If $E(X) = 4,5$ for a particular situation, what is $E(6X)$ equal to for this situation?

9. If $E(X^2) = 3$ for a particular situation, what is $4+2E(X^2)$ equal to for this situation?

10. If $E(X^2) = 3$ for a particular situation, what is $E(4+2X^2)$ equal to for this situation?

M_{37}: The Variance of a Random Variable Lecture 3

Objectives:

A. Students must know that the terms σ^2, Var(X) and $E(X-\mu)^2$ are all equivalent terms for the variance of a population.

B. Students must be able to calculate th variance of a sample (S^2) and the variance of a population (σ^2)

C. Students must be able to distinguish between the symbols used to disignate the variance (σ^2) and the standard deviation (σ) of a population.

D. Students must be able to apply the theoretical concepts of the mean and variance of a population (or the expected value and variance of a random variable) to practical situations.

Questions:

1. The variance of a population is represented by the symbol σ^2. What quantity or what parameter does the symbol σ represent for a population?

2. If you are given a set of data as follows:

x_i	f_i
10	3
4	3
6	2
9	1

and told that $\bar{x} = 7$, what is S_x equal to?

3. The variance of a population is represented by the symbol σ^2. In what other form do we represent or abbreviate the 'variance of a population'?

4. Consider a situation where a coin is spun twice. X, the random variable, is equal to the number of heads which occur and $\mu=1$. What is σ^2 equal to?
5. Consider a situation where the standard deviation $\sigma=2$ for a population. What is $E(X-\mu)^2$ equal to?
6. Consider a population where $E(X^2) = 7$ and $E(X) = 2$. What is Var(X) equal to for this population?
7. Consider the same situation as the one described in question 4. What is $E(X^2)$ equal to for this situation?
8. Say that a lottery is to be held. 1000 tickets are sold at R3 each. Three prizes are offered. The first prize is R500; the second prize is R300 and the third prize is R200. X, a random variable,is equal to the prize money. When a ticket is drawn it is replaced so that there are always 1000 tickets in the draw. What is the expected value of X,(E(X)),for this situation?
9. If you purchase one 3R ticket in the situation described in question 8, what is your expected gain or loss in this venture?
10. What is the variance of X,(Var(X)) equal to for the situation described in question 8?

M_{38}: The Binomial Probability Distribution Lecture 4

Objectives:

A. The students must be able to state the parameters of a binomial distribution when they are given the notation B(n;p) or b(x;n;p).

B. The students must be able to find binomial probabilities for different values of the random variable X in a practically stated problem.

C. The students must be able to identify and calculate the values of p, q, E(X), μ and σ^2 for a Bernoulli process.

Questions:

1. Consider a Bernoulli process where the probability of success p is equal to 0,8. What is the value of q?
2. What is E(X) equal to for a Bernoulli process?
3. In a particular clothing manufacturing company 75% of the employees are women and 25% are men. X is a Bernoulli random variable that takes on the value of 1 when an employee randomly selected is a woman and a value of 0 when a man is selected. What is μ equal to in this case?
4. Consider the same situation as the one described in question 3. What is σ^2 equal to in this case?
5. A particular Binomial distribution is specified as B(15;0,7). What is the probability of failure, q, equal to in this case?
6. Consider a certain university where 60% of the students are men. A random sample of 20 students is taken from the entire student population. The random variable X is equal to the number of men in the sample. What is the probability of obtaining 18 men in this sample?

7. What is the probability of obtaining at least 10 women in the sample described in question 6?

8. What is the probability of obtaining between 13 and 15 men inclusive in the sample described in question 6?

M_{39}: Properties of the Binomial Distribution Lecture 5

Objectives:

A. Students will be able to calculate the mean (μ or E(X)) and the variance (σ^2 or Var(X)) of a Binomial probability distribution when applied to a practical problem.

B. Students will be able to calculate the mean (μ or E(X)) and the variance (σ^2 or Var(X)) and the expected value of functions of X of a binomial probability distribution.

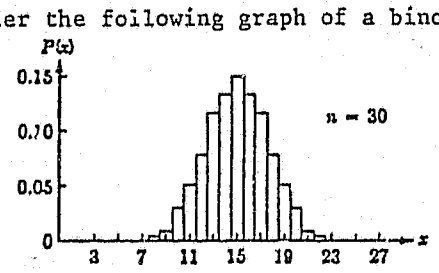
C. Students will be able to state the mean and variance of a Bernoulli and binomial distribution in terms of n, p and q (where n is the number of trials, p the probability of success and q the probability of failure).

D. Students will be able to identify and distinguish between the values of p and q when presented with graphs of different binomial distributions.

Questions:

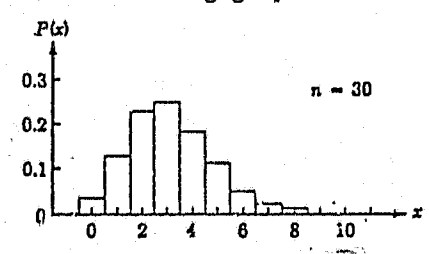
1. In a Bernoulli process n=1 and p+q = 1. What is the mean μ equal to for a Bernoulli process?
2. In a Bernoulli process n=1 and p+q = 1. What is the variance σ^2 equal to for a Bernoulli process?
3. A random variable X is equal to the number of heads which occur when a fair coin is spun 50 times. Thus the probability of success p=0,5 and the number of trials n=50. What is E(X) for this situation?
4. What is Var(X) equal to for the situation described in question 3?

5. Consider the following graph of a binomial distribution where n=30:



What is the value of p for this situation?

6. Consider the following graph of a binomial distribution where n=30:



Is p=q for this situation?

7. Consider the binomial distribution b(x;10;0,6). What is the expected value of X+3 equal to for this situation (E(X+3))?

8. Consider the binomial distribution b(x;10;0,6). What is the expected value of 3X+2 equal to for this situation (E(3X+2))?

9. Say that 60% of the residents of Johannesburg favour cinemas opening on a Sunday. Let X be equal to a binomial random variable which represents the number of people who favour cinemas opening on a Sunday. If a random sample of 20 residents is taken in Johannesburg, what is E(X) or the mean equal to for this sample?

10. What is the variance of X (Var(X)) equal to for the sample considered in question 9?

M_{40}: The Poisson Probability Distribution Lecture 6

Objectives:

A. The students will be able to state the formula for a Poisson distribution.

B. Given the formula for a Poisson distribution ($\frac{\mu^x . e^{-\mu}}{x!}$) and the values of μ and x, the students must be able to evaluate the formula, using a calculator.

C. Given the mean μ for a Poisson distribution, the students will be able to find the probability of a certain number of events occuring, with the aid of Poisson distribution tables.

D. Students will be able to calculate the difference between the mean μ and the variance σ^2 for a discrete distribution.

E. The students will be able to compare the arithmetic computation involved in evaluating ${}^nC_x p^x q^{n-x}$ when

(i) n is small (<25) and $p \simeq 0,5$

(ii) n is large (>25) and $p \simeq 0,1$

F. Students will be able to solve basic problems related to practical situations by applying the Poisson probability distribution.

Questions:

1. Consider a binomial distribution where n=3 and p=0,5. We would use the term ${}^3C_2 p^2 q^1$ in order to determine the probability of 2 successes occuring. Calculate the value of ${}^3C_2 p^2 q^1$ without using binomial tables.

2. Consider a binomial distribution where n=100 and p=0,1. If we wanted to calculate the probability of 32 successes occuring (P(X)=32), would we use this term: $\frac{100!}{32!68!} (0,1)^{32} (0,9)^{68}$?

3. Say we would like to find the probability of X being equal to a particular value in a Poisson distribution. The formula we would use to find this is: $\frac{(?)^x.e^{-(?)}}{x!}$. What is the name of the quantity which should be in the brackets where we have placed a question mark?

4. Consider a discrete distribution where n=100 and p=0,01. What is the value of $\mu-\sigma^2$ equal to for this situation.

5. Consider a Poisson probability distribution where μ=1,3. The formula for a Poisson distribution is $P(X=x) = \frac{\mu^x.e^{-\mu}}{x!}$. Use your calculator to find the probability that X will be equal to 3 (P(X=3)).

6. Consider a Poisson probability distribution where μ=2,5. The probability that X=6 can be found from the expression $\frac{(2,5)^6.e^{-2,5}}{6!}$. Use your calculator to evaluate this expression correct to 4 decimal places.

7. Consider a Poisson distribution where μ=3. Use your tables to find the probability that X=2.

8. Consider a Poisson distribution where μ=3. Use your tables to find the probability that $X \leq 2$.

9. Say that the average number of customers who appear at a bank teller's window is one per minute. Use the Poisson distribution where μ=1 to find the probability of four or more customers arriving at the teller's window in any particular minute.

Appendix II

Analyses and Instructional Changes

The instructional sequences and related questions which failed to attain the acceptance level of 70% are analysed in this appendix. The instructional changes that are made are viewed in relation to the results obtained by groups 1,2 and 3. Each question is analysed separately and in the following manner:

Objective: The objective which the students are expected to attain is stated.

Question: The question which tests the attainment of the objective is stated. A reference is given in the left margin so that the position of the question on the video cassette may be located. An asterisk refers to the original instruction and a plus to the changed or edited instruction.

Validity: The number of content specialists who rated the question as valid is given. All ratings are stated out of five as there were five content specialists. Thus 4/5 indicates that four out of the five content specialists rated the question as valid.

Instruction: brief description of the relevant instructional sequence is given. A video cassette reference is also given to facilitate the location of the sequence for viewing.

Result: The result obtained by group 1 for this question is stated followed by the result obtained by group 2 in brackets. Both groups viewed the identical instruction. Therefore 62% (58%) indicates 62% of the students in group 1 and 58% of those in group 2 answered the question correctly. This notation is also sometimes used in later discussion.

Analysis: An attempt is made to identify the area which is responsible for rendering the instruction ineffective. Once this has been done, the instructional change which is made to the television lecture in an attempt to eradicate this area is described.

Result: The result obtained by group 3, who viewed the 'improved' or changed instruction is given.

Conclusion: A final comment is made, regarding the effectiveness of the instructional change.

Lecture 1 - M_{35} Question 6

Objective: Students must be able to calculate the probability of a particular random event occuring in a sample space and be able to interpret notations such as P(X=3) and f(2) in these calculations.

*338 +335 Question: Consider a situation where a coin is spun three times. The random variable X is defined to be equal to the number of heads which occur. What is P(X=4) equal to?

Validity: 5/5.

*210 +211 Instruction: The students are shown the sample space with real coins which is generated when spinning a coin . They then write down this sample space, assigning probabilities of occurrence to the different random events. They are then shown an alternate way of writing this same sample space.

Result: 9% (2%) answered the question correctly.

Analysis: It was felt that the very low percentage of correct answers indicated that the students did not know where to begin to tackle the question. In other words they did not know how to generate a sample space for three spins of a coin as the instruction referred only to the case of two spins. The objective was that students are able to calculate probabilities, not generate sample spaces.

It was decided to add to the instruction, immediately after the question was asked, a description of how to generate a sample space for 3 spins of a coin.

Result: 25% answered the question correctly.

Conclusion: The change in the instruction made a considerable improvement to the result. However 25% is still a long way below the acceptance level of 70%. Thus further analysis is required. The objective should be re-examined and perhaps broken down into a few objectives, each with its own instructional sequence and question.

Lecture 1 - M_{35} Question 7

Objective: Students must be able to calculate the probability of a particular random event occurring in a sample space and be able to interpret notations such as P(X=3) and f(2) in these calculations.

*361 +394 Question: Consider a situation where a coin is spun four times. The random variable X is defined to be equal to the number of heads which occur. What is P(X=4) equal to?

Validity: 5/5

*210 +211 Instruction: Question 7 followed directly on from Question 6 as both questions were related to the same objective. The Instructional sequence is therefore identical to that for Question 6.

Result: 28% (18%) answered the question correctly.

Analysis: As the instructional sequence leading to Questions 6 and 7 was identical, the analysis is identical to that for question 6.

There is however a 19% (16%) increase in the number of correct answers. This is difficult to account for. A possible reason is that the students have had a little appropriate practive and thus improve on their answers the second time around.

The supposed improvement to the instruction remains the same as that for Question 6 with the addition that immediately after the question is asked, the students are reminded to extend the sample space which was developed for them immediately after Question 6 was asked. This is to obviate the difficulty in constructing a sample space which is not the objective in this situation.

Result: 55% answered the question correctly.

Conclusion: There is a marked improvement from 28% (18%) to 55%. This however is still below the acceptance level of 70%. The suggestion in the conclusion of question 6 is reinforced. If improved results are to be attained, the objective should be broken down into two or three parts.

A point to note is that a better result was obtained for this question than for question 6 although both were based on the same instruction. This appears to support the learning principle put forward in the analysis of the previous question that with some appropriate practive, the students improve their answers.

Lecture 1 - M_{35} Question 8

Objective: Students must be able to calculate the probability of a particular random event occurring in a sample space and be able to interpret notations such as P(X=3) and f(2) in these calculations.

*477
+508 Question: Consider the same situation as the one described in question 6. What is f(3) equal to?

Validity: 4/5

*210
+211 Instruction: Question 6,7 and 8 are all related to the same objective. Thus the instructional sequence leading to question 8 is the same as that for questions 6 and 7.

In question 8 a new notation is used when asking a similar question to the two previous questions. The following instructional sequence is used after question 7 to explain the new notation.

*379
+417 A reminder is given to the students concerning functions as encountered in school algebra. The concept of a function is then related to a probabilistic situation. The new notation for defining the probability of an event is then introduced. The students are then shown a probability distribution for the case of spinning a coin twice, using this new notation

Result: 33% (27%) answered the question correctly.

Analysis: It was felt that the problem area in the instructional sequence was identical to that for questions 6 and 7: the students were not adequately shown how to generate a sample space and thus could not proceed to the next step, which was to find the probability of an event taking place in the sample space. They could not therefore achieve the set objective.

The instructional changes effected for Question 6 and 7 should thus be adequate for this question, along with a small addition immediately after the question is asked, reminding the students to generate a sample space similar to the one they had used previously.

Result: 63% answered the question correctly.

Conclusion: This result almost attains the acceptance level of 70% even though no significant changes were made in the instruction for this question. This reinforces the hypothesis that the problem lay in the inability of students to generate sample spaces prior to question 6.

It appears that the learning principle of appropriate practice is revealed here as questions 6,7 and 8 are of a similar type and based on the same objective, and the results improved progressively for both the original instruction (9% (2%), 28% (18%), 33% (27%)) and the improved instruction (25%, 55%, 63%).

Lecture 1 - M_{35} Question 10

Objective: Students must be able to distinguish between a discrete probability distribution (probability function) and a continuous probability distribution (probability density function).

*579 +605 Question: If we are considering a probability density function, would this be a feasible statement to make: P(X=2) = 2?

Validity: 5/5

*496 +526 Instruction: Students are shown a table which outlines the differences between discrete and continuous probability distributions. To illustrate the differences further they are shown a graph of a discrete probability distribution followed by a question. They are then shown a graph of a continuous probability distribution.

Result: 66% answered the question correctly.

Analysis: When analysing the instruction preceding this question, it was decided to change the phrasing of the question. The wording could have caused unnecessary confusion. Thus the question was changed. This question is therefore rendered invalid for the experiment as the results obtained for different questions cannot be compared. This question was deleted from the experiment.

Lecture 2 - M_{36} Question 3

Objective: Students must be able to calculate the mean of a sample and the mean of a population.

*197
+229 Question: If you are given a set of data as follows:

x_i	f_i
10	3
4	3
6	2
9	1, what is $\bar{x}$ in this case?

Validity: 5/5

*172
+168 Instruction: Two formulae are reviewed, which were previously used by the students to find the mean of a sample. They are then asked question three.

Result: 18% (22%) answered the question correctly.

Analysis: Although the students had previously used both formulae to calculate the mean of a sample, they had either been unable to identify which of the two formulae to use in this instance or been unable to apply the formula to the data.

In an attempt to improve this instructional sequence it was decided to demonstrate the use of the correct formula on a very simple set of data. This was inserted immediately before question 3.

Result: 87% answered the question correctly.

Conclusion: The vast majority of the students were capable of achieving the objective, but only if their memories were refreshed.

Lecture 2 - M_{36} Question 4

Objective: Students must be able to distinguish between the relative frequency of occurrence $(\frac{f}{n})$ of an event (where the number of events n is finite) and the probability of occurrence $(\frac{f}{n})$ of an event (when the number of events n is infinite).

*341 +367 Question: Consider the expression $\lim_{n\to\infty} \frac{f}{n}$. What quantity or what concept does this expression define?

Validity: 5/5.

*231 +261 Instruction: The students are shown how to find the mean $(\bar{x})$ of a sample, using a formula for the mean which is arranged in a specific manner. Once this example has been completed, the formula is re-examined.

*272 +300 The specific manner in which the formula was arranged is pointed out as being significant. A segment of the formula $(\frac{f}{n})$ becomes the probability of an event occurring, if n tends towards infinity. This fact is now used in order to derive an expression for the expected value of a random variable.

Result: 26% (28%) answered correctly.

Analysis: The definition for the probability of an event occurring had been given to the students in a previous lecture. Therefore this question actually relates to the recall of knowledge. However, the students are reminded of this definition during the present instruction. This definition also plays a central role in the formula which is derived during this instructional sequence. Therefore the students should have been able to answer this question from what they derived from this sequence; the result however indicates that this was not the case.

It was decided that this lack of achievement was due to deficient instruction. Thus the entire instructional sequence which develops the formula for the expected value of a random variable was edited out and replaced with what was considered to be an improved explanation.

Result: 35% answered the question correctly.

Conclusion: This result is only a slight improvement and is still far short of the required level. Therefore the fault possibly does not lie with the instruction. The question was rated as fully valid. Thus the objective was reviewed.

It was decided that this objective is possibly out of place with this instructional sequence. The development of the sequence does not lead logically to this question. Therefore the question should perhaps be inserted at a different stage of the instruction - immediately after the students are reminded of the pertinent definition.

Lecture 2 - M_{36} Question 5

Objective: Students must know that the mean of a population and the expected value of a random variable are equivalent.

*358 +384 Question: The mean of a population $\mu=E(X)$. How do we read the expression $E(X)$?

Validity: 5/5.

*270 +299 Instruction: The students are shown how an equation for the mean of a population can be derived from the equation for the mean of a sample by considering the mean of a population to be the expected value of X, a random variable.

Result: 31% answered the question correctly

Analysis: It was considered that the form the answer to this question had to take was not suitable for recording efficiently and analysing its correctness on the Cybercom system. Therefore the question was changed so as to accommodate the Cybercom system. This question was deleted from the experiment because it now became a different question and comparisons between the answers from two different questions cannot be made.

Lecture 2 - M_{36} Question 6

Objective: Students must be able to calculate the expected value of: i) a random variable and ii) functions of a random variable.

*438 +467 Question: Consider a situation where a coin is spun twice. X, a random variable, is equal to the number of heads which occur. What is E(X) equal to?

Validity: 5/5.

*375 +406 Instruction: The students are shown how to find the expected value E(X) of a random variable X, for a population. The example which is chosen is that of throwing a die. They are shown how to draw up the probability distribution and how to apply the formula for E(X).

Result: 43% (27%) answered the question correctly.

Analysis: A result of 43% is considered too low. The instruction preceding the question however appeared to be adequate. A closer examination of the skills that are required to answer the question was made. These fell into two areas: i) the drawing up of a probability distribution which identifies the values of X and their corresponding probabilities of occurrence and ii) the application of the formula for E(X).

As the objective is concerned with the skills required in (ii), it was decided to illustrate (i) for the students in an attempt to improve the instruction. This was inserted immediately after question 6 was asked.

Result: 86% answered the question correctly.

Conclusion: The improved instruction has yielded a result well within the acceptable limit. This indicates that the students are capable of finding E(X) for a random variable X, but are not capable of generating probability distributions for X. More attention should be given to this aspect if further improvements to the instruction are to be made.

Lecture 2 - M_{36} Question 7

Objective: Students must be able to evaluate the expected value of (i) a random variable and ii) functions of a random variable.

Question: Consider the same situation as the one described in question
*498 +538 6. What is $E(X^2)$ equal to?

Validity: 5/5.

*461 +503 Instruction: The students are shown how the formula for E(X) applies to functions of X. A practical example is then demonstrated where $E(X^2)$ is found for the situation of throwing a die.

Result: 12% (6%) answered the question correctly.

Analysis: The analysis is identical to that for question 6. It was felt that the instruction was adequate but that two skills were needed to answer the question: i) the ability to draw up a probability distribution for values of the function of X and their corresponding probabilities of occurrence. ii) The application of the formula for the expected value of functions of X.

Again, as the skills in (ii) are the ones referred to in the objective, (i) was to be completed for the students. This was inserted immediately after question 7 was asked.

Result: 84% answered correctly.

Conclusion : This vastly improved result reinforces the conclusions reached for question 6 as question 7 is a duplication of that situation.

Lecture 2 - M_{36} Question 10

Objective: Students must be able to evaluate the expected value of i) a random variable and ii) functions of a random variable.

*603 +649 Question: If $E(X^2)=3$ for a particular situation, what is $E(4+2X)$ equal to for this situation?

Validity: 5/5

*571 +617 Instruction: The students are shown two properties of expected values, namely that i) the expected value of the sum of two random variables is equal to the sum of the expected value of each random variable and ii) the expected value of a constant added to a constant multiplied by a random variable is equal to the first constant added to the second constant which is multiplied by the expected value of the random variable. Two questions are then asked on these two properties, questions 9 and 10.

Result: 68% (55%) answered correctly.

Analysis: The result obtained by group 1 (68%) was very close to the acceptance level. It was therefore decided to leave the instructional sequence unchanged and observe if groups 2 and 3 scored a better result.

Result: 56% answered correctly.

Conclusion: All three groups failed to achieve the level of acceptance. This instrucional sequence is probably inadequate and should be reviewed in order to improve the attainment of the objective. If a similar example were to be worked out for the students before the question is asked, a larger percentage should achieve the objective.

Lecture 3 - M_{37} Question 2

Objective: Students must be able to calculate the variance of a sample (S^2) and the variance of a population (σ^2).

*142 +173 Question : If you are given a set of data as follows:

x_i	f_i
10	3
4	3
6	2
9	1

and told that $\bar{x}=7$, what is S_x^2 equal to?

Validity: 5/5

*101 +97 Instruction: Two formulae which were previously used by the students to calculate the variance of a sample are reviewed. They are then told to use the second formula to be reviewed in order to answer question 2.

Result: 5% (5%) answered the question correctly.

Analysis: It was assumed that the students were familiar with the application of this formula to a set of data. Obviously this assumption was incorrect.

It was therefore decided, in an attempt to improve the instructional sequence, to illustrate the use of this formula when applied to a simple data set. This was inserted immediately before the question was asked.

Result: 36% answered the question correctly.

Conclusion: The illustrative example which was inserted resulted in a substantial increase in the number of correct answers (31%). The overall number of correct answers is still however far too low (36%). A problem area has therefore been identified.

The students had obviously not reached a sufficient level of competence in applying the relevant formula to sets of data in order to find the variance of a sample. Thus their earlier instruction should be suitably changed. Alternately more time could be devoted to this area in this instructional sequence before the question is asked.

Lecture 3 - M_{37} Question 4

Objective: Students must be able to calculate the variance of a sample (S^2) and the variance of a population (σ^2).

*321 +344 Question: Consider a situation where a coin is spun twice. X, the random variable, is equal to the number of heads which occur and $\mu=1$. What is σ^2 equal to?

Validity: 5/5

*281 +305 Instruction: The students are shown how to calculate the variance of a population by considering the situation where a die is thrown. They are then asked to find the variance for the situation described in question 4.

Result: 24% (35%) answered the question correctly.

Analysis: The instruction was considered adequate although the result was far from satisfactory.

This lead to a closer examination of the skills required to answer the question correctly. Again, it was found that the necessary skills could be separated into two distinct operations. The first involved the generation of a probability distribution for the random variable X and the second the application of the formula for the calculation of the variance of X.

As the objective referred only to the second operation it was proposed that an improved result could be obtained if the first operation was completed for the students. This was done immediately after the question was asked.

Result: 75% answered the question correctly.

Conclusion: This result is within the acceptance level. A skill or operation has been identified which was not part of this specific objective. It was also in this area that the students were experiencing difficulty. Further practice at generating probability distributions is needed by the students.

Lecture 3 - M_{37} Question 5

Objective: Students must know that the terms σ^2, Var(X) and $E(X-\mu)^2$ are all equivalent terms for the variance of a population.

*378 +419 Question: Consider a situation where the standard deviation $\sigma=2$ for a population. What is $E(X-\mu)^2$ equal to?

Validity: 5/5

*351 +390 Instruction: The two formulae which have been developed so far for E(X) and Var(X) are compared. From the comparison it is shown that the variance of X, Var(X) can also be written as $E(X-\mu)^2$.

Result: 43% (49%) answered the question correctly.

Analysis: Although they were shown two equivalent forms of writing the variance of a population, these two forms were never related to the most often used form, σ^2. It was felt that this was a deficiency in the instruction.

As a result, it was decided to stress that σ^2, Var(X) and $E(X-\mu)^2$ are all equivalent forms, immediately after the first two equivalent forms had been stated.

Result: 76% answered correctly.

Conclusion: The deficiency in the instruction had been correctly identified as a relatively small change in the instruction resulted in a significantly better result.

Lecture 3 - M_{37} Question 7

Objective: Students must be able to calculate the variance of a sample (S^2) and the variance of a population (σ^2).

*487 +523 Question: Consider the same situation as the one described in question 4. What is $E(X^2)$ equal to for this situation?

Validity: 3/5

*403 +442 Instruction: The students are shown how to calculate the variance for the situation of throwing a die by using the formula $Var(X)=E(X^2)-[E(X)]^2$. They are reminded that they have already seen that $E(X)=3,5$. They are then shown how to find $E(X^2)$ for this situation and finally that Var(X) is equal to $\frac{35}{12}$.

Result: 19% (31%) answered the question correctly.

Analysis: In order to answer the question the students had to first of all generate the probability distribution for X^2 and then find $E(X^2)$. The objective only refers to calculating the variance of X which involves evaluating $E(X^2)$.

As was the case in question 4, the probability distribution for X^2 was generated for the students immediately after the question was asked, in an attempt to improve the instruction.

Result: 64% answered correctly.

Conclusion: The stating of the probability distribution increased the correct responses by 45% (33%) which is a significant improvement. However the acceptance level has not yet been attained. This would indicate that more time should be spent in explaining the application of the new formula for Var(X) and the calculation of terms such as $E(X)$.

Lecture 3 - M_{37} Question 8
Question 9
Question 10

All three questions are related to the same objective and instructional sequence. As it was decided to leave the sequence unchanged, the questions will be discussed together as a group.

Objective: Students must be able to apply the theoretical concepts of the mean and variance of a population (or the expected value and variance of a random variable) to practical situations.

*538 +580 Question 8: Say that a lottery is to be held. 1000 tickets are sold at R3 each. Three prizes are offered. The first prize is R500; the second prize is R300 and the third prize is R200. X, a random variable, is equal to the prize money. When a ticket is drawn it is replaced so that there are always 1000 tickets in the draw. What is the expected value of X, [E(X)], for this situation?

*562 +603 Question 9: If you purchase one R3 ticket in the situation described in question 8, what is your expected loss or gain in this venture?

*580 +620 Question 10: What is the variance of X, Var(X) , equal to for the situation described in question 8?

Validity: 5/5 for all three questions.

*504 +547 Instruction: A practical situation is stated where the theory developed in this lecture can be applied. A probability distribution is drawn up for the students and then three questions based on the situation are asked.

Result: Question 8: 56% (72%) answered correctly.
Question 9: 14% (18%) answered correctly.
Question 10: 10% (4%) answered correctly

Analysis: Although all three of these questions failed to achieve the minimum acceptance level, the related instructional sequences were left unchanged. They are all related to the same objective, namely that students must be able to apply the theoretical concepts of the mean and variance of a population to practical situations.

The students were mostly unsuccessful in obtaining this objective, as the scores for the three questions indicate. However, it was felt that in order to apply theoretical concepts to practical situations, students must first have mastered the theoretical concepts. This does not appear to be the case in many instances during the earlier parts of this lecture.

Therefore, as it was felt that improvements made to earlier sequences of the instruction would more readily facilitate the achievement of these prerequisite objectives, this objective would also be more readily attained, and the achievement levels might well be within the acceptance level without any instructional changes.

Result: Question 8: 67% answered correctly.
Question 9: 0% answered correctly.
Question 10: 3% answered correctly.

Conclusion: Although earlier sequences have shown a larger percentage of students acquiring mastery of subordinate objectives, these results are, in the main, worse. Therefore the instructional sequence should be changed. This change should include the solution of problem of a similar type to that given to the students to solve.

Lecture 4 - M_{38} Question 2

Objective: The students must be able to identify and calculate the values of p, q, E(X), μ and σ^2 for a Bernoulli process.

*183 +178 Question: What is E(X) equal to for a Bernoulli process?

Validity: 5/5

*164 +157 Instruction: The students are shown the probability distribution for a Bernoulli process and are then asked to find the expected value of X for a Bernoulli process.

Result: 40% (44%) answered the question correctly.

Analysis: In previous lectures it was thought that students experienced difficulty in generating probability distributions in order to find E(X). This cannot be the case here as they are shown the probability distribution. Some students may not however realize that this is the appropriate probability distribution to use.

A second source of difficulty may be that students cannot recall the formula for E(X) from the previous lecture.

As the objective does not concern itself with generating probability distributions or with the recall of a formula, but with the calculation of E(X) for a Bernoulli process, the students were told that the distribution which they had just seen for a Bernoulli process was the appropriate distribution and they were reminded verbally of the formula for E(X). This was all inserted immediately after the question was asked.

Result: 71% answered the question correctly.

Conclusion: This result is now within the acceptance limit, although only just. Further improvements to the instruction should be sought in order to increase the percentage of correct answers.

Lecture 4 - M_{38} Question 3

Objective: The students must be able to identify and calculate the values of p, q, E(X), μ and σ^2 for a Bernoulli process.

*276 Question: In a particular clothing manufacturing company 75% of the em-
+276 ployees are women and 25% are men. X is a Bernoulli random variable that takes on the value of 1 when an employee randomly selected is a woman and a value of 0 when a man is selected. What is μ equal to in this case?

Validity: 4/5

*215 Instruction: The students are shown that the variance for a Bernoulli
+211 process is equal to pq. While this is being done they use the answer to question 2, namely that E(X)=p.

Result: 41% (43%) answered correctly.

Analysis: In the instruction they are shown two properties of a Bernoulli process, namely. E(X)=p and Var(X)=pq. When question 3 was asked, the symbol μ was used. Some students might not associate μ and E(X) as being equivalent, which they should have done from a previous lecture.

In an attempt to improve the instruction and highlight these two properties it was decided to insert a summary of them at the end of the instructional sequence, immeidately before question 2 was asked. The equivalent symbols of μ and σ^2 were used in place of E(X) and Var(X) respectively in the summary.

Secondly, many students might have difficulty in identifying the values of p and q in the question. Although this was part of the objective it was decided that the focus should be on the calculation and thus they are given the values of p and q. The students were expected to perform two skills: identify p and q and calculate μ. By giving them the value of p and q, only one skill is now to be performed.

Result: 80% answered the question correctly.

Conclusion: The instruction is now considered adequate to enable the vast majority of the students to achieve part of the objective. In order to enable students to achieve the entire objective, it should be broken down into two parts, where in the first part students are required to identify p and q for different Bernoulli processes. Only then should the present instruction be given and question 3 subsequently asked.

Lecture 4 - M_{38} Question 4

Objective: The students must be able to identify and calculate the values of p, q, E(X), μ and σ^2 for a Bernoulli process.

*303 +309 Question: Consider the same situation as the one described in question 3. What is σ^2 equal to in this case?

Validity: 4/5

*215 +211 Instruction: The instruction is identical to that for question 3.

Result: 48% (26%) answered the question correctly.

Analysis: The analysis of the instruction is identical to that for question 3 as both questions stem from the same instruction. Thus the instructional changes made for question 3 should increase the number of correct answers for question 4.

In addition, the students are reminded to use the property of the variance for a Bernoulli process which had just been developed.

Result: 55% answered the question correctly.

Conclusion: This result does not show much of an improvement and it is still well below the acceptance level. Thus the instruction preceding questions 3 and 4 is probably not quite appropriate, even with the 'improvements'.

Further improvements would probably include some examples of finding μ and σ^2 for a Bernoulli process before the students are asked to perform a similar calculation.

Lecture 4 - M_{38} Question 6

Objective: The students must be able to find binomial probabilities for different values of the random variable X in a practically stated problem.

*487 +492 Question: Consider a certain university where 80% of the students are men. A random sample of 20 students is taken from the entire student population. The random variable X is equal to the number of men in the sample. What is the probability of obtaining 18 men in this sample?

Validity: 5/5

*454 +460 Instruction: The situation as stated in question 6 is presented to the students and then a particular question based on this situation is completed for them; namely the probability is found of there being 15 women in the sample.

Result: 5% (18%) answered the question correctly.

Analysis: The skills required to answer this question were closely examined and it was found that there were two, possibly unrelated, skills involved.

The first required that the students identify which value or values of the random variable X are involved with a particular value of p.

The second required the student to find the probabilities associated with these values of X and p.

The objective as it is stated refers to the second of these skills. Thus, in order to improve the instruction, it was decided to perform the first skill for the students on the chalkboard.

Result: 55% answered correctly.

Conclusion: This result represents a 50% (37%) improvement which is substantial. However the result still falls below the level of acceptance and further improvements should take place.

The problem which is worked out for the students before the question is asked could be explain more fully with every step in the calculation being shown. The situation at the moment is that a few steps are missed as it was assumed the students would follow these rather elementary calculations and procedures. Obviously this was not the case.

Lecture 4 - M_{38} Question 7

Objective: The students must be able to find binomial probabilities for different values of the random variable X in a practically stated problem.

*524 +548 Question: What is the probability of obtaining at least 10 women in the sample described in question 6 ?

Validity: 5/5

*504 +530 Instruction: The instruction is similar to that for question 6. Before the question is asked a similar question is worked out for the students.

Result: 4% (39%) answered the question correctly.

Analysis: The analyses is identical to that described in question 6. Thus in order to improve matters, the first skill is again performed +553 for the students.

Result: 61% answered correctly.

Conclusion: Although the result represents an improvement of 57% (22%) and is only 9% below the level of acceptance, the remarks made in the conclusion of question 6 also apply here.

Lecture 4 - M_{38} Question 8

Objective: The students must be able to find binomial probabilities for different values of the random variable X in a practically stated problem.

*560 +613 Question: What is the probability of obtaining between 13 and 15 men inclusive in the sample described in question 6?

Validity: 5/5

Instruction: The instruction is similar to that for question 6. Before the question is asked a similar question is worked out for the students.

Result: 32% (57%) answered correctly.

Analysis: The analysis is identical to that described in question 6. Once again, the first skill described in the analysis of question 6 is +620 performed for the students on the chalkboard.

Result: 66% answered the question correctly.

Conclusion: Although the result is very close to the acceptance level of 70% the conclusion reached in question 6 still applies here.

It is worth noting that although the instruction remained essentially the same for questions 6, 7 and 8 for both the original instruction and then for the 'improved' instruction, the results mostly improved [5% (18%) → 4% (39%) → 32% (57%) and 55% → 61% 66%] progressively as the students worked from question 6 through to question 8. This appears to support the proposal made in the conclusion of question 6 that further improvements can be effected if the problems completed before each question are more thoroughly treated - the students' results tended to improve as they became more familiar with that type of problem. Therefore the learning principle of appropriate practice should be applied more stringently.

Lecture 5 - M_{39} Question 1

Objective: Students will be able to state the mean and variance of a Bernoulli and Binomial distribution in terms of n, p and q (where n is the number of trials, p the probability of success and q the probability of failure).

*52 +61 Question: In a Bernoulli process n=1 and p+q=1. What is the mean μ equal to for a Bernoulli process?

Validity: 5/5

*31 +29 Instruction: The first question is based on a formula derived in the previous lecture. Thus after a brief preamble which describes what the present lecture is going to cover, the first question is asked.

Result: 45% (3%) answered the question correctly.

Analysis: Two formula which were developed in the previous lecture are essential to the development of this lecture. Thus instead of assuming that the students would be familiar with them, at the beginning of this lecture the students are asked to state the two formulae in questions 1 and 2.

The result obtained indicates conclusively that the students are not familiar with the first formula. Thus it was decided to insert a brief review of these two formulae in the preamble to this lecture and to observe the effect.

Result: 36% answered correctly.

Conclusion: The brief review did not improve the initial result, although there is a large fluctuation in the initial result. This highlights the variation in recall between groups. Thus it appears that a brief review is not sufficient. As the development of this lecture is based, to a large extent, on the students being familiar with these two formulae, a more substantial review would be in order.

Lecture 5 - M_{39} Question 2

Objective: Students will be able to state the mean and variance of a Bernoulli and Binomial distribution in terms of n, p and q (where n is the number of trials, p the probability of success and q the probability of failure).

*227
+236 Question: In a Bernoulli process n=1 and p+q=1. What is the variance σ^2 equal to for a Bernoulli process?

Validity: 4/5

*71
+81 Instruction: Question 1 was based on the fact that $\mu=p$ for a Bernoulli process. This relationship is now developed to apply to a Binomial process or distribution and it is shown that $\mu=np$ for this situation. The next instructional sequence is to derive a relationship for the variance (σ^2) of a Binomial distribution. As this is based on the formula for the variance of a Bernoulli process (which was dealt with in the previous lecture), this is the question which is now asked as question 2.

Result: 43% (39%) answered the question correctly.

Analysis: This is identical to the analysis outlined in question 1.

Result: 88% answered the question correctly.

Conclusion: A substantial increase in the number of correct answers has resulted. This result is well within the level of acceptance, whereas the result for question 1 was not. This is possibly accounted for by the fact that between questions 1 and 2 the instructional sequence includes expressions of a similar type to those used in the previous lecture which act as stimuli or reminders to the students.

The proposed further improvement in the conclusion of question 1 is still deemed to be necessary.

Lecture 5 - M_{39} Question 4

Objective: Students will be able to calculate the mean (μ or E(X)) and the variance (σ^2 or Var(X)) of a Binomial probability distribution when applied to a practical problem.

*395 +420 Question: A random variable X is equal to the number of heads which occur when a fair coin is spun 50 times. Thus the probability of success p=0,5 and the number of trials n=50. What is Var(X) equal to for this situation?

Validity: 3/5

*250 +258 Instruction: The students are shown a second property of the binomial distribution, namely that the variance σ^2=npq. They are then asked a question (question 3) on the first property they were shown and then a question on the second property (question 4).

Result; 52% (62%) answered the question correctly.

Analysis: Questions 3 and 4 are based on similar instruction. Question 3 was very well answered, with 76% (82%) of the students submitting the correct answer.

The instruction was therefore possibly adequate. The lower result in this question was possibly due to the fact that a large amount of ground was covered before questions 3 and 4 were asked. The students had possibly assimilated the first property but were still absorbing the second. It was decided therefore to present a summary of these two important properties of the binomial distribution at the end of the instructional se-

+378 quence, immediately before questions 3 and 4 were asked.

Result: 84% answered correctly.

Conclusion: This result is well within the acceptance level and the instructional sequence is far more satisfactory since the summary was introduced.

Lecture 5 - M_{39} Question 10

Objective: Students will be able to calculate the mean [μ or E(X)] and the variance [σ^2 or Var(X)] of a Binomial probability distribution when applied to a practical problem.

*654 +673 Question: What is the variance of X [(Var(X)] equal to for the sample considered in question 9?

Validity: 5/5

*629 +649 Instruction: A practical situation is stated where the binomial distribution can be employed. This fact is made known to the students and two questions are posed concerning this situation; questions 9 and 10.

Result: 65% (79%) answered correctly.

Analysis: The score obtained for this question by group 1 was very close to the acceptance level. Coupled to this is the belief that earlier parts of the instruction in this lecture have been improved. Therefore it was decided to leave this instructional sequence unchaged.

Result: 80% answered correctly.

Conclusion: As both Groups 2 and 3 scored above the acceptance level it appears that the instruction was adequate and the decision to leave it unchanged is vindicated.

Author Young D A
Name of thesis Computer-aided evaluationt of television instruction in a Tertiary-level introductory statistics course 01153

PUBLISHER:
University of the Witwatersrand, Johannesburg
©2013

LEGAL NOTICES:

Copyright Notice: All materials on the University of the Witwatersrand, Johannesburg Library website are protected by South African copyright law and may not be distributed, transmitted, displayed, or otherwise published in any format, without the prior written permission of the copyright owner.

Disclaimer and Terms of Use: Provided that you maintain all copyright and other notices contained therein, you may download material (one machine readable copy and one print copy per page) for your personal and/or educational non-commercial use only.

The University of the Witwatersrand, Johannesburg, is not responsible for any errors or omissions and excludes any and all liability for any errors in or omissions from the information on the Library website.