A STUDY OF ORDERING MECHANISMS FOR GENERATING R6 AND SAI DISTRIBUTIONS FROM CRITERIA-BASED ASSESSMENT DATA

A THESIS
SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR THE DEGREE OF MASTER OF EDUCATION
AUSTRALIAN CATHOLIC UNIVERSITY

Brett Rangiira BSc., Grad. Dip. Teach.
Department of Professional Development
Australian Catholic University
Queensland Division
Brisbane 4000
Australia

April 1996

ABSTRACT

This study explores the issue of generating ordered distributions of Queensland senior secondary school students using criteria-based assessment data. Criteria-based assessment data are multidimensional, yet ordered distributions are unidimensional since order is a property of univariate data. Ordering multivariate data requires the data to be transformed onto a univariate scale and such transformations are called here ordering mechanisms.

The ordered distributions obtained from the criteria-based assessment data are used for certification purposes and tertiary entrance selection purposes. Importance is placed on these distributions, yet there seems to be little study of the ordering mechanisms to yield these distributions from multidimensional criteria-based assessment data. This study seeks to address this gap.

The purpose of this research is to explore the issue of generating macro-function assessment data from micro-function assessment data. More specifically, this research considers the issue of ordering multivariate educational assessment data for percentage-based assessment systems, and in particular, the nature of the ordering mechanisms for ordering systems within such assessment systems.

This study is focused by the following research question:

RESEARCH QUESTION:

Is there a mathematical system which:

(1) orders all points in a multidimensional percentage-based assessment space,

(2) models preferences within a multidimensional percentage-based assessment space,

(3) can be used to construct an independent ordering system for a percentage-based assessment system?

The existence of a mathematical system for modelling ordering mechanisms has been questioned. This study shows that a mathematical system with properties which suggest its applicability for developing ordering mechanisms in assessment systems for certain Board subjects, does exist.

DECLARATION

I certify that this thesis does not contain any material previously submitted for a degree at any University; and to the best of my knowledge and belief it does not contain any material previously submitted or written by another person except where due reference is made in the text.

Brett Rangiira

ACKNOWLEDGEMENTS

I would like to thank my wife Nicole for her support and encouragement throughout this research. It was her belief that this study would finally come together which allowed it to continue through times of greatest difficulty. It would not have been possible without her.

I would also like to thank my family and friends for the significant contribution which they have made in supporting me throughout this study.

Finally, I wish to thank my supervisor, Dr. Dennis McLaughlin, for his support and critical advice, and for guiding this project through to completion.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

CHAPTER ONE
THE RESEARCH ISSUE

1.1 Introduction

This study investigates the issue of generating ordered distributions of students using educational assessment data. More specifically, this study focuses on one aspect of the assessment procedures for students graduating from Queensland senior secondary school education (years 11 and 12), namely, the requirement to order assessment data generated within the criteria-based assessment paradigm.

The term 'criteria-based assessment data' (CBAD) will be used here to refer to the assessment data generated within the criteria-based assessment paradigm. Criteria-based assessment was introduced into Queensland secondary schools following recommendations from the Scott Report (Scott et al., 1978), referred to as ROSBA, an acronym derived from part of the title of the report, i.e., *The Review Of School Based Assessment.* At the time, criteria-based assessment was considered a panacea for the problems which arose from the use of norm-based assessment within the context of school-based assessment (Scott et al., 1978), (to be addressed in section 2.3). Under norm-based assessment, grades were integers ranging from 1 to 7, with 7 the highest, and these were defined statistically as regions of fixed area under a standard normal probability distribution. The raw data required to facilitate this grading system is numerical univariate (i.e. single score) interval scale data. Data of this nature are necessary for grading using a normal distribution. Such data easily facilitate grading since they are inherently ordered, but foster a competitive agenda within the classroom, causing tension and undermining quality learning (Scott et al., 1978, pp. 5-7). Perhaps more importantly, however, univariate interval scale data do not provide sufficient feedback to students to facilitate the learning process. To tell a student '22 out of 53 marks' does little to explain to the student how and why they are performing in such a way, and what such a performance suggests about their level of understanding, knowledge, and the implications for future learning. To facilitate learning, assessment should provide multidimensional, descriptive data referenced to the requirements of the task environment, which is the context of the learning activity and performance. Hence, when criteria-based assessment replaced norm-based assessment, one of the major implications was the kind of assessment data generated under the new assessment paradigm. CBAD are multidimensional where the indicators and descriptors of achievement within each dimension (called criterion (plural: criteria) in criteria-based assessment) are a combination of numbers, letter codes and verbal descriptors.

One of the strengths of criteria-based assessment is that the CBAD are multivariate and not restricted to numerical scales. The data are considered to be of a form which facilitate the learning process. Also, the grading system consists of five Level of Achievement (LOA) categories and does not require overall achievement to be expressed as a position on a univariate interval scale. One of the weaknesses of criteria-based assessment, however, is that because CBAD are multivariate, CBAD are not ordered since 'order properties ...exist only in one dimension' (Kendall, quoted in Barnett, 1976, p. 318). The requirement to generate ordered distributions from CBAD raises the important question of how ordered distributions are to be derived from data which are not ordered. This question appears to have been largely overlooked; this author found no acknowledgement of the issue of ordering multivariate CBAD in the Board of Senior Secondary School Studies (BSSSS, or the Board) documentation (BSSSS, 1992, 1993a), or in reports pertaining to the generation of the ordered distributions of CBAD (Viviani, 1990). There is reason to believe that this issue has not been investigated (Sadler, 1988, p. 10).

A formulation of the research question is in section 1.5. The assessment milieu which gives rise to this question is complex and designed in accordance with the Queensland senior secondary school education system. Prior to a formulation of the question, the context of educational assessment in Queensland needs to be clarified. Section 1.2 will introduce the main elements of the assessment procedures relevant to the research question. Sections 1.3 and 1.4 outline the main conceptual framework within which the research question is defined. This framework will seek to provide a broader context for the Queensland situation, linking it to broader scholarship in educational assessment. Section 1.6 will outline the design of the research, which is further developed in the methodology chapter, chapter 3. Section 1.7

presents an outline of the thesis.

1.2 Assessment in Queensland senior secondary school education

This section seeks to explain the elements of the assessment procedures in Queensland senior secondary school education. These elements will provide a framework which will be used to derive the research question.

The Board is the curriculum authority for Queensland senior secondary school education. The Board oversees syllabus development, and is responsible for the accreditation of school work programs and the certification of student assessment. The Board has existed in one form or another since the Education Act of 1964 provided for a Board of Junior Secondary School Studies and a Board of Senior Secondary School Studies. The two boards were replaced by a Board of Secondary School Studies in the early 1970's following recommendations by the Radford Report (Radford, 1970) and this single Board was changed back to the Board of Senior Secondary School Studies following the Education (Senior Secondary School Studies) Act 1988.

The relationship between the Board and schools is described as a 'partnership...for the purpose of satisfying the claims and expectations of society about accountability in the education system as well as the specific needs of individual students' (BSSSS, 1993a, p. 1). Together, the Board and schools cooperate in offering subjects to students in senior secondary school education. To explain the nature of this partnership it is important to outline first the kinds of subjects offered in Queensland senior secondary school education.

There are four broad categories of subjects offered in Queensland senior secondary school education: Board subjects, Board registered subjects, Recorded subjects and School subjects. It is the assessment procedures for Board subjects which are relevant to this study. By way of definition:

> Board subjects are those subjects for which:
> - syllabuses have been approved by the Board
> - work programs are subject to Board accreditation procedures

- assessment of student achievement are subject to the full certification procedures of the Board
- time tabled school time devoted to study and assessment of the subject is a minimum of 55 hours per semester (BSSSS, 1993a, p. 29).

Board subject syllabi are developed by Subject Advisory Committees (SACs) and outline the material to be covered by the subject, define the criteria to be used in assessment within the subject, and provide descriptions of the five LOA in terms of criteria and standards. A school wishing to offer a Board subject must develop a work program from the syllabus detailing the specific way the subject is to be taught and assessed in the school. Syllabi provide a degree of flexibility which allows for variation in the work programs for a Board subject between schools. The Board accreditation procedures serve to accredit the school work program thereby allowing the school to offer that subject to its students.

It is the school which designs the learning experiences and develops and administers the assessment for a Board subject, but it is the Board's responsibility to award certificates to students stating achievement in Board subjects. Schools award LOAs but the Board awards the certificates on which these are recorded. This is one of the major partnerships which exist between the Board and Schools. Since assessment in Queensland is school-based, it is necessary to ensure that assessment is comparable from school to school. To ensure comparability across Board subject assessment, the Board has established the process of certification.

> Certification is the process by which school decision making on standards of student achievement is verified by matching student performance with stated criteria and standards in the school's accredited work program (BSSSS, 1993a, p. 3).

Certification occurs in two stages. The first is called monitoring which is:

> the process by which panels consider the decisions made by schools on standards of achievement in Board subjects after year 11 so that advice can be given to schools to assist and reassure them on their judgements (BSSSS, 1993a, p. 55).

The second stage is called reviewing and is:

> the process near the end of the course in October and November when District and State Review Panels consider submissions of student work and other supporting documents and advise schools on their proposals for awarding Levels of Achievement

to students about to exit from the course (BSSSS, 1993a, p. 63).

This study investigates the issue of generating ordered distributions of students from CBAD. It is the Board which requires these ordered distributions and there are two of them in all. The first is a distribution of the entire student cohort for a Board subject throughout the 5 criteria-based LOA. This distribution is required for the reviewing process, and is a graphical display of the student cohort on a ladder of 50 rungs, 10 rungs per LOA. (The five LOA are Very High Achievement (VHA), High Achievement (HA), Sound Achievement (SA), Limited Achievement (LA) and Very Limited Achievement (VLA). The rungs on the R6 are labelled using the LOA category and the rung number within that category, e.g. SA5 is rung 5 in the SA category.) This distribution appears on a form called Form R6 (appendix A) which is one of a series of forms required for reviewing. The term Form R6, or simply R6, will be used here to refer to the distribution of a student cohort on this scale.

The R6 is used in the following manner. A school will prepare a submission consisting of the accredited work program, samples of student work selected from the cohort for students in each LOA, and the R6 distribution showing the distribution of the cohort together with the selected students positioned on the ladder and coded to match each folio in the submission. This submission is sent to a District Review Panel consisting of teachers of that Board subject from local schools. Review panellists check the school's assessment instruments against the work program and decide if the placements of the sample student folios on the R6 are appropriate. At the end of the school year, after the final school-based assessment has been conducted, the school will negotiate Exit Levels of Achievement with the Board via the Review Panels. These negotiations are facilitated using the R6.

The second ordered distribution of the student cohort required by the Board are a set of numbers called Subject Achievement Indicators (SAIs) which range from 200 to 400. SAIs indicate the relative overall achievement between students within a Board subject within a school and are used by the Board to calculate a State wide rank order of students from which student Overall Positions (OPs) are derived. OPs are numbers between 1 and 25, 1 the highest, and are used as the primary selection instrument for entrance into tertiary education.

They are defined as regions along a continuum of scaled overall achievement expressed on an interval scale. The SAIs provided by the school are scaled using the results from a State wide examination called the Queensland Core Skills Test (QCST) to yield the scaled overall achievement. The other numbers used for tertiary entrance are Field Positions (FPs) which are a set of five numbers (per student) ranging from 1 (the highest) to 10.

The SAIs were introduced into Queensland in 1992 following recommendations from the Viviani Report (Viviani, 1990). SAIs replaced SSAs, numbers ranging from 0 to 99 which were used to calculate TE (Tertiary Entrance) scores (BSSSS, 1988) prior to the introduction of the OP FP system following the Viviani Report. The TE score was itself introduced into the Queensland secondary education system in 1973 following the transition from Public Examinations to school-based assessment, as recommended in the Radford Report. The reasons for the change to OPs and FPs are provided in the Viviani Report (1990, p. 17, p. 22) but will be discussed no further here.

Both the R6 and the SAI ordered distributions are to be derived from the CBAD generated throughout the course of study for a given cohort for a given Board subject within a given school. Hence, there are three types of data, CBAD, R6 and SAI, and R6 and SAI are generated from CBAD (Figure 1.1)

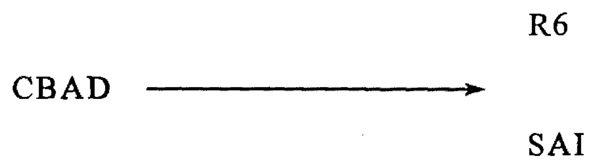


Figure 1.1: R6 and SAI are derived from CBAD

There are three categories of student cohort for a given Board subject, defined according to the size of the cohort: small group, intermediate group and large group (BSSSS, 1993a, p. 67). An R6 is required for all three groups, but SAIs are determined only for large groups. For small and intermediate groups, the data used for calculate OPs are taken directly off the R6 distribution. This relationship between the R6 and the SAIs raises some questions. Are the SAIs intended to be a transformation of information from the R6 scale, or are they to be obtained independently from the R6 and somehow directly from the CBAD?

Literature provided by the Board to assist schools in determining SAIs suggests that SAIs are to be derived from the R6 distribution.

> Exit levels of achievement must be established first. SAIs are then derived from the levels of achievement by means of a rank order of students within each level of achievement (BSSSS, 1992, p. 5).

The R6, however, is itself not mentioned directly although it is implied by 'rank order of students within each level of achievement'. If the SAIs are intended to be obtained from the R6 distribution, then they would be calculated directly from the R6. A problem arises since 'SAIs are [to be] decided rather than calculated' (BSSSS, 1992, p. 3). What is meant by 'decision' and 'calculation' is not explained, nor is the reason why a decision is preferred over a calculation. A procedure for deciding SAIs is provided in this literature, but reflection on the process outlined reveals problems. This procedure is outlined below.

- Place top student at one end of a line
- Place last student at other end of the line
- Select any other student
- Find two students, one ranked above (nearest), the other below (nearest) this student
- Place this student relative to these two (if there is no information to suggest otherwise - place this student in the middle of the other two)
- Given this process for three students, extend to 30 or 300

(BSSSS, 1992, p. 6).

Consider 5 students numbered from 1 to 5. Students 1 and 2 are top and bottom respectively. These students are placed at either end of a line. Suppose students 3, 4 and 5 are nearest each other on some overall achievement continuum (perhaps the R6). Line 5 in the procedure for determining SAIs requires 3 to be placed on the straight line relative to 4 and 5. Student 4 and 5, however, are not positioned on the straight line at this stage. In placing student 3 on the line, the decision should be one of determining where student 3 lies with respect to the top and bottom student in the cohort which are probably not the students nearest student 3 unless there are only 3 students in the cohort. Where should student 3 be placed relative to the top and bottom students? The procedure does not answer this question.

The R6 and SAIs are, by their very nature, ordered distributions, hence unidimensional. The SAI scale is an interval scale, but it is uncertain whether the R6 is interval in nature. The rungs within an LOA may constitute an interval scale, but this does not mean, for example, that a jump from VLA10 (rung 10 VLA) to LA1 is the same as the jump from rung SA10 to HA1. It may be 'easier' to go from a VLA to a LA than from an SA to a HA. Although there are tests to determine the nature of measurement scales (Nuthmann, 1994, pp.4-16), this issue is not be pursued here. What is known, however, is that if the R6 is not interval throughout, then the mapping of data from the R6 to the SAI interval scale is not a permissible transformation (Krantz, Luce, Suppes, & Tversky, 1971, pp. 8-12).

To conclude this section a summary of the elements of the assessment procedures in Queensland senior secondary school education is provided. The diagram in Figure 1.2 shows nine elements of the assessment procedure. A cohort of students in a Board subject in a given school are assessed regularly throughout years 11 and 12 (element 1). The assessment is designed by the teachers within that subject department within the school. Student responses to the various assessment tasks, e.g. exams, projects, assignments etc., together with the teacher judgements made on the student responses, are collected for each student and retained in a student assessment folio (element 2). Teacher judgements are then collated to yield CBAD profiles of student achievement for each student (element 3).

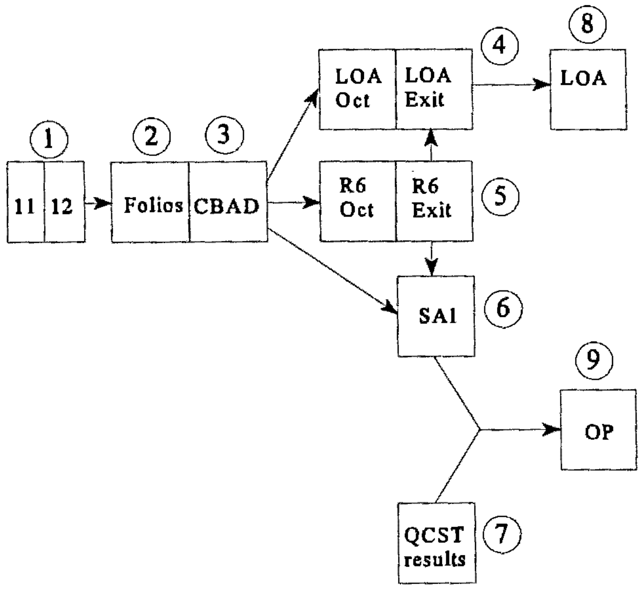


Figure 1.2: Elements of the assessment procedures for Queensland senior secondary school education

Toward the end of year 12 (September/October) Exit LOA (element 4) are forecast for each student in the cohort using the available CBAD. The cohort of students is placed on the R6 (element 5) and a submission is sent for certification by the district panel. Certification involves teachers from outside the school making judgements about the school's assessment program and decisions. Negotiation occurs between the panel and the school regarding the positioning of the students on the R6, and if agreement cannot be reached the submission is sent to the state panel for that Board subject for further deliberation. By the end of the school year, the distribution of the cohort on the R6, hence the number of students in each LOA category and the Exit LOA (element 8) for each student has been finalised.

The SAIs (element 6) are determined after the R6 is finalised. The SAIs for the cohort of students in each Board subject within the school are then sent to the Board. This data is scaled using the QCST results to yield OPs (element 9).

This study is concerned with the ranking of students within Board subjects to yield R6 and SAI distributions. The issue of calculating OPs from SAIs and QCST results is not addressed in any detail.

The R6, SAI and CBAD are elements unique to the Queensland assessment procedures. To link this study to the broader field of scholarship in educational assessment, it is necessary to locate the R6, SAI and CBAD within a more general framework. Such a framework will now be identified.

1.3 Micro- and macro-functions of assessment

This section will introduce an important element of the conceptual framework for this study. The two major elements of this framework are 'micro-functions of assessment' and 'macro-functions of assessment'. These concepts are in turn supported by the notion of the micro- and macro-dimension of the education system (senior secondary school education system). The education system can be conceived as a system of nested systems (Biggs & Moore, 1993, p. 450): classrooms are nested inside schools which are inturn nested within a broader State

Department structure. The term 'micro-dimension of the education system' will be reserved for the system of a Board subject within a school. The term 'macro-dimension of the education system' will be reserved for the system of bodies which lie beyond schools and function as a centralised authority on curriculum and assessment issues. The Board, together with the District and State Review Panels which act as an arm of the Board for certification, constitute the macro-dimension of the education system with which this study is concerned. The micro- and macro-dimension of the education system are nested components of the one system.

A schematic diagram of the micro- and macro-dimension of the education system is shown in Figure 1.3. The District and State Review Panels are assumed by the Board. The two cohorts of students at any time throughout a school year (year 11 and year 12) together with the classes for each of these cohorts are shown in the micro-dimension. In this model, schools link the micro-dimension to the macro-dimension.

Both the micro- and macro-dimensions of the education system serve a range of functions. The micro-dimension is primarily concerned with the education of cohorts of students in a Board subject within a school. One of the major roles of the micro-dimension therefore is to foster teaching and learning, and for this study, fostering teaching and learning will be called the micro-functions of the education system, i.e. the functions of the micro-dimension of the education system. The macro-dimension on the other hand, whilst it is concerned with teaching and learning, is also responsible for certification and the calculation of OPs for tertiary entrance. For this study, certification and the calculation of OPs for tertiary entrance will be called the macro-functions of the education system.

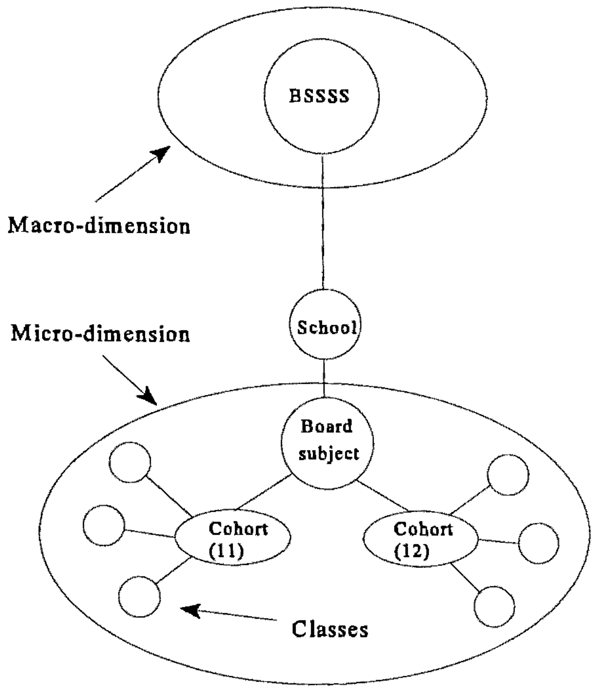


Figure 1.3: Micro- and macro-dimensions of the education system

Educational assessment is a key element for both the micro- and macro-functions of the education system. For the micro-functions of the education system, assessment provides feedback to teachers and students, informing teachers and students about the learning process and thereby directing future learning (Findlay, 1987). For the macro-functions of the education system, assessment provides data which facilitate both reviewing (the R6) and the calculation of OPs (the SAIs). The term micro-functions of assessment (section 2.2.1) will refer to the role which assessment has in facilitating the micro-functions of the education system. The micro-functions of assessment are to foster and promote student learning. Similarly, the term macro-functions of assessment (section 2.2.2) will refer to the role which assessment has in facilitating the macro-functions of the education system. Macro-functions of assessment foster State wide ordering processes.

The kind of assessment data generated for use within the micro-dimension of the education system will be called micro-function assessment data (MiFAD), and the assessment data generated for use within the macro-dimension of the education system will be called macro-function assessment data (MaFAD). MiFAD should serve to promote the micro-functions of assessment. The transition to criteria-based assessment from norm-based assessment was in part due to the nature of the assessment data generated under norm-based assessment. Norm-based assessment data was MiFAD which was inconsistent with the micro-functions of assessment because it failed to provide a profile of information about student achievement and resulted in negative backwash. (The term backwash refers to the impact which assessment has on student learning (Biggs & Moore, 1993, p. 391).) Docking (1987, p. 10) articulates the problems arising from the use of norm-based assessment as MiFAD.

> Most of the destructive effects arise from the vagueness and relativity of systems that determine standards by reference to norms rather than by reference to competencies, and include student alienation, stress and anxiety, excessive competitiveness, extrinsic motivation, curriculum and assessment invalidity, meaningless grades, loss of accountability, restricted individualism, ambiguity in goals and standards, test 'gamesmanship', narrowed curriculum and restricted expectations.

CBAD, on the other hand, is considered to be consistent with the micro-functions of assessment. The Scott Report (Scott et al., 1978) advocates criteria-based assessment over norm-based assessment by arguing that criteria-based assessment creates a more desirable learning environment within the classroom. This view that criteria-based assessment, or characteristics of assessment systems similar to criteria-based assessment, fosters classroom learning is supported elsewhere (Docking, 1986, 1987; Sadler, 1986a, 1987).

Just as MiFAD should promote the micro-functions of assessment, so MaFAD should promote the macro-functions of assessment. MaFAD is unidimensional interval scale data (at least for the SAIs) and this is necessary for the mathematical scaling procedures which are required to determine the OPs. Because norm-based assessment data are unidimensional interval scale data, such data easily facilitate the macro-functions of assessment. CBAD, however, does not. CBAD are multidimensional, and not necessarily interval scale numerical data across the criteria. Because of this, raw CBAD are not of a form which can be used as MiFAD. Under norm-based assessment, MaFAD was obtained via a simple mathematical transformation of data from one unidimensional interval scale to another. Criteria-based assessment gives rise to an entirely different situation. Because CBAD are multidimensional, MaFAD cannot be obtained as a mathematical transformation of data from one interval scale to another because CBAD is not expressed as positions on an interval scale. This issue will be elaborated on in section 1.4 and 1.5.

To conclude this section, a diagram (Figure 1.4) showing the micro- and macro-functions of assessment has been provided. The term 'assessment system' is used in this diagram and throughout this study, to refer to the set of assessment procedures, policies, structures, concepts and processes for a Board subject within a given school. It is necessary to restrict the term to this specificity because assessment not only differs from Board subject to Board subject, but also from school to school since assessment in Queensland is school-based.

The diagram shows two cycles, one for the micro-functions of assessment and one for the macro-functions of assessment. The cyclic nature represents a feedback mechanism. Assessment is an integral aspect of the curriculum and does not function as a separate and external element. It has been suggested that a dynamic dialogue is established between the micro-dimension and macro-dimension of the assessment system and the assessment used to judge it.

> Test scores, rather than playing the role of passive indicator variables for the state of the system, become the currency of feedback within an adapting educational system. The system adjusts its curricular and instructional practices, and students adjust their learning strategies and goals, to maximise the scores on the tests used to evaluate educational outcomes (Frederiksen & Collins, 1989, p. 27).

Having established this framework, the next stage is to take a closer look at the assessment system at the centre of the diagram in Figure 1.4. This is where the specific focus of this study lies.

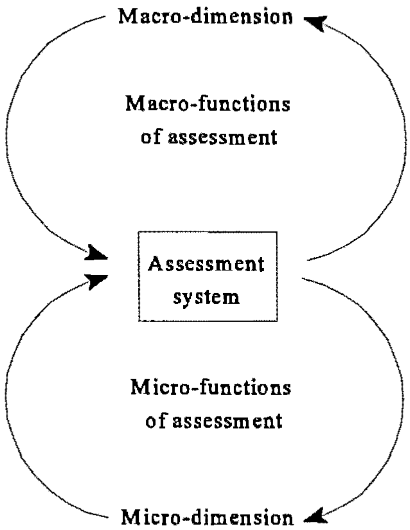


Figure 1.4: Micro- and macro-functions of assessment

1.4 The assessment system

The assessment system generates MiFAD and transforms MiFAD to MaFAD. Under the current assessment procedures in Queensland, MiFAD is CBAD and MaFAD is R6 and SAI. Because the MaFAD represents an ordering of the student cohort, the transformation of MiFAD to MaFAD represents an ordering of mulidimensional data. The term 'ordering system' (OS) will refer to the transformation of MiFAD to MaFAD. The ordering system is shown (Figure 1.5) as the interface between the micro- and macro-functions of assessment.

The issue for this study is located with the OS in Figure 1.5. Another term, called 'ordering mechanism' (OM) will now be introduced and used to model an OS. For certain types of data, the OS is a one step mathematical transformation from one interval scale to another. For other types of data there are a number of different possible approaches to structuring an OS and these may involve more than one step. This requires some explanation. Suppose CBAD is 1, R6 is 2 and SAI is 3. These different OS are shown in Figure 1.6. The U in Figure 1.6 (c) is an intermediatory unidimensional (hence U) interval scale.

Each arrow represents an OM which is a transformation of data from one form to another. An OS consists of a set of OM.

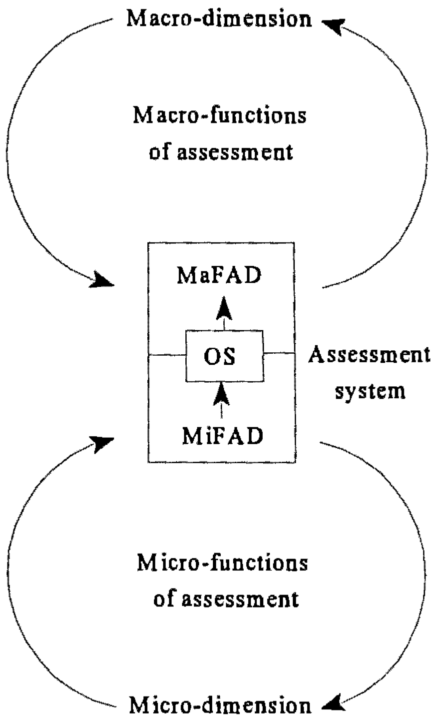


Figure 1.5: The ordering system

The two OS structures considered in this study are those in Figure 1.6 (a) and (c). The OS in Figure (a) will be called a 'dependent OS' because the SAI distribution is obtained from the R6 distribution, hence the SAI is dependent on the construct of the R6. Also, the initial

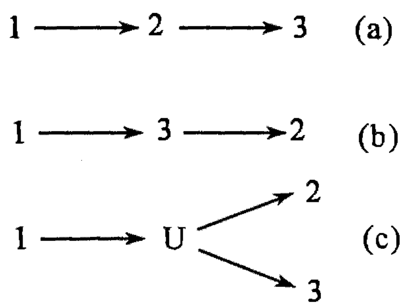


Figure 1.6: Three different ordering systems

ordering of the data occurs directly on the R6 scale, hence the fundamental order derived from CBAD occurs on the R6 and is dependent on the R6. The OS in Figure (c) will be called an 'independent OS' because both the R6 and the SAI are obtained from a scale which is independent of the constructs of each. This study is primarily concerned with the independent OS.

The independent OS consists of three different types of OM. These are OMu, OMr6 and OMsai. These three OM are shown in Figure 1.7. This model provides a general structure for the independent OS, but clearly there are many different types of independent OS because there are a variety of different OMu, OMr6 and OMsai.

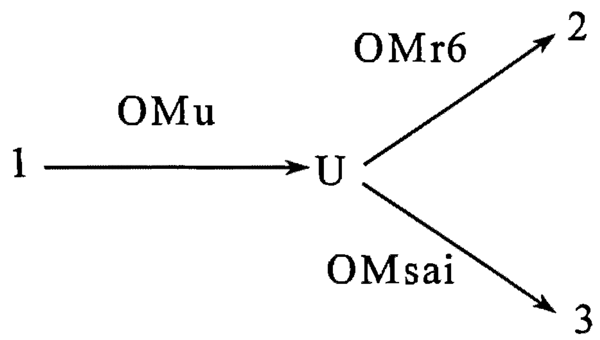


Figure 1.7: The independent ordering system

Consider the following example of independent OS. A teacher has a bundle of student assessment information for each student in a particular cohort. To generate the R6 and SAI distribution, the teacher decides to throw the bundles down a flight of stairs with 25 steps. In this way the teacher obtains a unidimensional distribution ordered on a scale from 1 to 25. The act of throwing the bundles down the stairs is an OMu where each bundle gets a number from 1 to 25. The teacher then decides to determine LOA as VLA 1-5, LA 6-10, SA 11-15, HA 16-20, VHA 21-25, reasoning that 'the heavier they are, the more ink used to comment, hence the more errors, hence the heavier the papers, hence the further they will fall'. Each step represents two rungs, and the teacher then flips a coin to decide on which of the two rungs each bundle should be placed: heads up a rung, tails down a rung. The SAIs are a transformation of the data from the 1 to 25 scale such that the bottom number gets 200 and the top gets 400.

The dubious independent OS in this example illustrates two important points. Firstly, OM are not necessarily mathematical operations, since throwing the bundles is more of a physical operation. Mathematical operations are only possible if the data are numerical. An OM could be a discernment process where teachers discuss a student profile at length and debate until agreement is reached about the position on a univariate interval scale. Secondly, OM represent an underlying set of principles, beliefs and constructs about overall achievement: 'the heavier they are, the more ink used to comment, hence the more errors, hence the heavier the papers, hence the further they will fall'.

Whilst an OMr6 and OMsai are of importance for the independent OS, for CBAD the fundamental ordering of the students occurs with OMu. Although the R6 and SAI distributions may quite different for different OMr6 and OMsai, the ordering of the candidates on U must be preserved, since OMr6 and OMsai represent a transformation of this ordering onto the R6 and SAI scales. Of primary interest in this study is the OMu for the independent OS.

Some Board subject syllabi, e.g. senior chemistry, provide school with the option of generating numerical data within each of the assessment criteria of the subject and allow schools to summarise achievement within each criteria as a percentage of marks obtained out of a total number of marks. The profile of student achievement for these subjects is reduced to a multivariate variable where the components of each dimension are percentages. For a subject with n criteria, achievement is represented as an n-tuple of percentages, e.g., for two criteria, a result would be of the form (x, y) where x and y are percentages. Assessment systems which generate data of this type will be called 'percentage-based assessment systems'.

One OMu for generating a distribution for percentage-based assessment systems involves obtaining the average of the percentages in each dimension. The practice of adding results across criteria to yield a single result, however, is itself an issue which has caused some debate in the literature. Averaging involves tradeoff and Hewitson (1988, p. 248) argues that 'tradeoffs compromise the logic of criterion referenced assessment' because 'a founding principle of the criterion referenced approach is that the criteria need to be defined as discrete qualities' (the term criterion referenced is not synonymous with criteria-based (section 2.3) but seems to be used synonymously by many authors). Arguments against aggregation create problems for criteria-based assessment. On the one hand it is argued that the process of aggregating results across criteria is educationally meaningless (Sadler, 1988; Docking, 1976) because a single result does not represent the detail of student performance, detail necessary for facilitating micro-functions of the education system (teaching and learning). On this basis it is invalid and should not occur. On the other hand, aggregation of one form or another is necessary for obtaining MaFAD since the MaFAD is an order, and order implies aggregation since it is a property of univariate data, not multivariate data. This dilemma is a fundamental issue at the interface of the micro- and macro-functions of assessment and one addressed by this study.

The position taken in this study is that it is necessary to combine the results of multivariate CBAD to yield the R6 and SAIs. Questions of the educational meaningfulness of aggregating results, although of fundamental importance, are not addressed here. What is being addressed here is the process of OMu, what the OMu represent in terms of decisions and the kinds of models which can be used as OMu. Herein lies the purpose of this research.

1.5 Purpose of the research

The purpose of this research is to explore the issue of generating macro-function assessment data (MaFAD, i.e., R6 and SAI distributions) from micro-function assessment data (MiFAD, i.e., CBAD). More specifically, this research considers the issue of ordering multivariate educational assessment data for percentage-based assessment systems, and in particular, the nature of the ordering mechanisms for ordering systems within such assessment systems.

The emphasis on criteria, standards and profiles of achievement in Queensland education, the acknowledgement of the destructive effects which the application of mathematical models (i.e. normal distributions) have on the curriculum, and the preference of multidimensional profiles over unidimensional grades and scores, seems to have left unaddressed the important issue of how to order multidimensional data, and the necessary unidimensional nature of ordered distributions. Furthermore, the requirement that the process of determining SAIs be decisions not calculations (Viviani, 1990, p. 133) raises important questions about the use of mathematical models for assessment processes. Do 'calculations' necessarily undermine quality decisions, or can decisions be modelled using mathematics and be calculations? This study addresses this issue and seeks to articulate the structure of ordering systems, and in particular, mathematical systems which can be used to design OMu for ordering systems.

Percentage-based assessment systems generate data which are able to be manipulated using mathematical models. If there are mathematical models to use as OMu, then these models must be shown to be modelling teachers decisions about the overall achievement represented by a profile of numbers. The existence of a mathematical system to generate models for this purpose, however, has been questioned (Sadler, 1988, p. 10).

> It has yet to be demonstrated whether a mathematical system can be developed which faithfully maps (i) the global achievement, (ii) allowable tradeoffs on different criteria and (iii) non-negotiable minima on some or all criteria, onto a final achievement scale which satisfactorily models teachers' carefully considered global judgements. There are reasons for believing that it is impossible with any form of additive composition rule.

The term 'decision structure' will be used to refer to teachers' judgement constructs about overall achievement, including preferences. The research question for this study is derived from the following broader question:

> Is there a mathematical system which is capable of representing decision structures as OM for independent OS within percentage-based assessment systems?

This question can be considered in three parts. The first part concerns the decision structures which underlie teachers' decisions about overall achievement for a profile of results. These decision structures function either implicitly or explicitly when teachers make these sorts of decisions. For this reason, these decision structures are constructs which lend themselves to empirical investigation. The second part concerns the mathematical system which is able to represent these decision structures as ordering mechanisms. Such a mathematical system would contain the building blocks for fashioning specific models to represent particular decision structures as ordering mechanisms. This part of the question is non-empirical. It seeks abstract structures which possess generic properties of decision structures, not specific properties, and can therefore occur separately to empirical study. The third part is the verification of such a mathematical system by using it to build ordering mechanisms within assessment systems. Like the first part of the question, this is also an empirical issue since the success of the model to fulfil this task is in the end judged by the teachers, students and assessment authorities which it serves.

This research is largely focussing on the second part of the question as outlined above. To do this, some generic properties of decision structures must be identified. Two will be identified here. The first is that decision structures must order all points in the assessment space (the term assessment space refers to the set of all possible values for the multivariate data). The second is that different decision structures reflect different preferences for points in the assessment space. For example, one decision structure may prefer the result (100, 0) over the result (50, 50) thereby ranking (100, 0) higher on the univariate scale of achievement than (50, 50). Another decision structure might yield the reverse.

This study is focused by the following research question:

RESEARCH QUESTION:

Is there a mathematical system which:

(1) orders all points in a multidimensional percentage-based assessment space,

(2) models preferences within a multidimensional percentage-based assessment space,

(3) can be used to construct an independent ordering system for a percentage-based assessment system?

This research question has emerged from this author's own experience with the assessment procedures outlined here. When faced with the task of generating R6 and SAI distributions from CBAD, it became apparent that there is no overall ordering of CBAD since CBAD are multivariate data. The important question of how the CBAD should be ordered seemed not to be addressed in the literature. Reflection and some initial investigation into this issue — motivated by the importance attributed to the R6 and SAI distributions and the need to explain to parents and students how they are derived — revealed a complex assessment milieu consisting of artefacts from prior legislation, reports and Board documents, e.g.: school-based assessment, criteria-based assessment, certification, R6 distribution, criteria and standards, profiles of achievement, Levels of Achievement, SAIs (SSAs and TE Scores) OPs and FPs. It was discovered that the transfer to criteria-based assessment in 1978 and the retention of the Order of Merit List introduced in 1973 from Radford recommendations introduced into the Queensland assessment procedures the problem of ordering multivariate data, a problem acknowledged by some (Sadler, 1988) but largely not addressed.

This study is an attempt to address the issue of ordering multivariate assessment data, an issue occurring at the interface of the micro- and macro-functions of assessment. This issue is embedded in the assessment processes which have evolved in Queensland senior secondary school education since the introduction of school-based assessment.

1.6 Design of the research

This study uses a research method called modelling and simulation, used largely in the discipline of operations research. A mathematical system is designed (modelled) and its properties are investigated (simulation). In this way, a mathematical system is developed and its properties are considered in light of the requirements outlined in the research question.

A detailed outline of the research design is provided in chapter 3. A brief outline of the research is provided below.

First, this research will develop a mathematical system to order all points in a multidimensional assessment space. A review of the literature found that the concept of ordering multivariate data is addressed in many different ways, largely in philosophy and mathematics. Many of the treatments of order, however, were found not to be appropriate for this study. Philosophical discussions of order (Lorand, 1992), Set Theory (Kamke, 1950) and Statistics (Barnett, 1976) addressed some of the fundamental issues regarding order, which are relevant to this study, but these discussions did not present a model of the nature which is sought after here (section 2.4). The literature which did address this issue (Keeny & Raffia, 1976) was found in 'decision theory'. There is a topic in decision theory called 'multiattribute utility functions' (section 2.4.4) and within this topic are found value functions which are used to model decisions using multiple attributes. This topic in decision theory provided the inspiration for the mathematical system developed here (section 4.2 and 4.3).

Second, this research will investigate the properties of the mathematical system developed in section 4.2 and 4.3 in light of the requirement to model preferences. This will involve computer simulation investigations.

Finally, this study will design an independent ordering system for a percentage-based assessment system and will generate R6 and SAI distributions for a cohort of senior chemistry students from Queensland school. The calculated distributions will then be compared with the R6 and SAI distributions determined by the school.

1.7 Outline of the thesis

This thesis consists of five chapters. This, the first chapter, has presented a general model of the ordering systems in Queensland school-based assessment systems and in doing so has attempted to explicate the context of this research from which the research question has been framed. It has located the research issue at the interface of the micro- and macro-functions of assessment. More specifically it has suggested that the requirement for assessment data to facilitate both the micro- and macro-functions of assessment introduces into school-based assessment systems the problem of ordering multivariate assessment data, and the need to articulate the structure of ordering systems.

The second chapter is a literature review and explores issues in the literature which are considered to be of most relevance to this study. Three areas from the literature have been researched: the functions of educational assessment (section 2.2), criteria-based assessment (section 2.3) and ordering multivariate data (section 2.4).

The third chapter outlines the method of research used in this study. The method is identified as modelling and simulation, a method used largely in the discipline of operations research. The application of this method to a problem arising in the field of education constitutes educational research since 'education is not itself a discipline' but rather

> *a field of study*, a locus containing phenomena...problems...and processes, which themselves constitute the raw material for inquiries of many kinds. The perspectives and procedures of many disciplines can be brought to bear on the questions arising from and inherent in education as a field of study. As each of these disciplinary perspectives is brought to bear on the field of education, it brings with it its own set of concepts, methods, and procedures, often modifying them to fit the phenomena or problems of education (Shulman, 1988, p. 5).

The fourth chapter contains the results and discussion of the research, and the fifth chapter is a synthesis and review of the study.

CHAPTER TWO
LITERATURE REVIEW

2.1 Purpose of research

The purpose of this research is to explore the issue of generating macro-function assessment data from micro-function assessment data. More specifically, this research considers the issue of ordering multivariate educational assessment data for percentage-based assessment systems, and in particular, the nature of the ordering mechanisms for ordering systems within such assessment systems.

This study is focused by the following research question:
RESEARCH QUESTION:

Is there a mathematical system which:

(1) orders all points in a multidimensional percentage-based assessment space,

(2) models preferences within a multidimensional percentage-based assessment space,

(3) can be used to construct an independent ordering system for a percentage-based assessment system?

The research question arises from the broad issue of generating rank order R6 and SAI distributions of students from CBAD. This arises at the interface of the type of data required to facilitate the micro- and macro-functions of assessment (section 2.2). Micro-functions of assessment facilitate classroom learning processes (section 2.2.1) by providing feedback to teachers and students. Teachers and students respond to such feedback (Meisels, 1989; Shepard, 1989) and behave as a dynamic system which 'adapts itself to the characteristics of the outcome measures' (Frederiksen & Collins, 1989, p. 27). Arising out of this is the requirement for assessment to be 'systemically valid'. Systemically valid assessment 'induces in the education system curricular and instructional changes that foster the development of the cognitive skills the test is designed to measure' (Frederiksen & Collins, 1989, p. 27). This implies the need for a non-competitive, task-referenced (as opposed to norm-referenced) assessment system, where assessment data can be interpreted by students in terms of the requirements of a given task domain. Herein lies the virtues of criteria-based assessment (sections 2.3). Micro-functions of assessment require assessment data to be multidimensional (multivariate) and referenced to absolute standards.

Macro-functions of assessment (section 2.2.2) facilitate state-wide certification and selection processes. The nature of these functions require that assessment data be ordered, and hence unidimensional, and referenced to both absolute standards (R6, section 1.2), and to the achievement of other students (SAIs, section 1.2). Within a school-based assessment system, MaFAD must be generated from MiFAD (section 1.3), a situation which in Queensland requires the ordering of mulitvariate criteria-based assessment data. Ordering multivariate data, however, is a complex issue (Barnett, 1976). Its complexity stems from the fact that 'order properties...exist only in one dimension' (Kendall, quoted in Barnett, 1976, p. 318) and multivariate data must be transformed onto a univariate scale for there to be order. The nature of these transformations, although discussed in the fields of measurement theory (Krantz et al., 1971) utility theory (Berger, 1985), and decision theory (Keeny & Raffia, 1976), seem to have received little attention within educational literature (Sadler, 1988).

The literature which has been reviewed for this study has been organised into three sections. The first section (section 2.2) examines the diverse range of functions of educational assessment. These functions operate at both micro- and macro-dimensions of the education system and on this basis can be broadly grouped into two categories: micro-functions (section 2.2.1) and macro-functions (section 2.2.2).

The second section (section 2.3) examines the origins and elements of the criteria-based assessment paradigm. It seeks to outline the nature of the multidimensional MiFAD which is generated within this paradigm. In particular, it will address the meaning of the terms criteria, standard, and level of achievement, and the implications of these concepts for assessment data.

The third section (section 2.4) considers the issue of ordering multivariate data. It will explore some different approaches to understanding the nature of order and will present the concepts from multiattribute utility theory which are used in this study to address the research question.

2.2 The functions of educational assessment.

The study of educational assessment probably should not begin with a search for a tight definition. The literature offers a variety of 'definitions', some of which appear to be contradictory.

> Assessment is placing an interpretation on measurement information (Smith & Lovat, 1991, p. 152).
>
> Assessment can be practised without any kind of measurement (Rowntree, 1977, p. 5).
>
> Assessment is concerned with assigning a mark, a rank, a grade, or some qualitative comment to measurement information (Smith & Lovat, 1991, p. 152).
>
> Assessment is not the same thing as grading or marking...Assessment can be descriptive without becoming judgemental (Rowntree, 1977, p. 5).
>
> The term assessment is interpreted broadly, so as to include any judgement (or appraisal, or evaluation) of a student's work or performance made by a teacher or other competent person, whether for purposes of improvement or certification (Sadler, 1987, p. 191).

Another starting point is to investigate the functions, or purposes, of educational assessment.

Educational assessment serves a broad range of functions, and some authors have grouped these functions into two broad categories (Scriven, 1967; Vandome (cited in Rowntree, 1977); Rowntree, 1977; Stiggins & Bridgeford, 1985; Docking, 1987). The first category consists of those functions which facilitate classroom based teaching and learning processes. Within the entire education system the classroom exists at the micro-dimension and for this reason these functions of assessment will be called 'micro-functions' (section 1.3). These micro-functions of assessment are elsewhere referred to as 'formative assessment' (Scriven, 1967), 'pedagogic assessment' (Vandome, cited in Rowntree, 1977), and 'teachers primary measurement needs' (Stiggins & Bridgeford, 1985).

The micro-function label (and macro-function, see below) is coined here to emphasise the idea that the two categories of assessment functions operate at different dimensions of the single entity called the education system. The distinction between 'large-scale' and 'small-scale' assessment functions is acknowledged by Docking (1987), but the idea of the education system as being a system of nested systems, i.e., classrooms nested within schools, schools nested within state departments of education, departments nested within society at large (Biggs, 1991, pp. 221-229; Biggs & Moore, 1993, p. 450), provides a particularly useful framework for locating these large-scale and small-scale assessment functions as functional aspects of a system.

This invites elaboration. A system has been defined as 'any entity, conceptual or physical, which consists of interdependent parts' (Ackoff, 1969, p. 332). The implication of systems theory for understanding educational structures and processes, and as a paradigm for generating curriculum theory, has been acknowledged by curriculum writers (Gough, 1989; Slaughter, 1989; Doll, 1989). Educational systems are essentially functional systems (or 'goal directed', or 'directively organised', (see Nagel (1969), for a discussion of the formalisation of functionalism), which means that there exists a teleological dimension to their existence, and consequently their structure. They serve a range of purposes, and these purposes are oriented at different dimensions of the system.

When viewed as a system of nested systems, there is a hierarchical dimension to the organisation of the component nested systems, hence micro- and macro-dimensions (section 1.3, Figure 1.2). The micro-dimensions are the classrooms and schools, the macro-dimensions are, for example, the Board of Senior Secondary School Studies, The Education Department, and the Tertiary Entrance Procedures Authority (TEPA). Furthermore, these micro- and macro-dimensions of the education system serve different purposes. Classrooms are primarily concerned with facilitating student learning. TEPA for example is concerned with selection of students into tertiary education.

The past 25 years have seen a gradual shift in the use of assessment paradigms within education systems from those which facilitate the macro-functions of assessment to those which are micro-function oriented (Judge, 1985). The transition from norm-based assessment to criteria-based assessment in the 1970's in Queensland is a prime example. Docking (1987, p. 11) summarises Judge (1985) who identifies trends in educational measurement which are moving towards criterion-referenced measurement (in the context used here by Docking criterion-referenced assessment can for all intents and purposes be considered to be criteria-based assessment). Some of the trends noted by Docking are in Table 2.1.

Table 2.1
Trends in educational measurement (as noted by Docking, 1987)

Past approaches	Future approaches
Summative	Formative
System Centred	Student Centred
Competitive	Competitive and Co-operative
External/Centralised	Internal/School-based
Norm-referenced	Criterion-referenced

The requirement for a single assessment paradigm to satisfy dual micro- (student centred) and macro-functions (system centred) of assessment generates some significant challenges.

> How can the system needs [macro-functions] and the teacher needs [micro-functions] be met simultaneously? One strategy for dealing with this problem is to establish dual systems of assessment, one a summative system for credentialling needs, and the other a formative system for teacher and student needs. Such an approach is doomed to fail because of the additional workload involved, and because most students will only take seriously the system from which their grades are determined.
>
> If dual systems are likely to exacerbate problems, a single system must be devised that will meet all needs. Most writers refer to criterion-referenced measurement as a means of meeting the needs of the students and the teachers, but this approach presents problems when it is required to interface with an institutional or state grading system (Docking, 1987, p. 11).

2.2.1 The micro-functions of assessment

The micro-functions of assessment are primarily concerned with the process of learning. For example, The Scott Report (Scott et al., 1978, p. 40) advocates five purposes of assessment, three of which are micro-functions:

- to enable the student to assess his/her learning and to provide information which can be used to correct deficiencies in his/her learning;
- to enable the teacher and students to modify teaching/learning programs to suite the

needs of the students;

- ultimately to provide the student and parents with information concerning the student's achievement and potential.

Elsewhere Docking (1987, p. 10) lists five micro-functions of assessment:

- Student achievement;
- Student diagnosis and readiness;
- Student motivation and guidance;
- Student feedback;
- Student-self awareness and self-assessment.

Rowntree (1977) acknowledges four:

- motivation of students (p. 21);
- feedback to students (p. 24);
- feedback to the teacher (p. 27);
- preparation for life (p. 28).

Micro-functions of assessment facilitate classroom teaching and learning processes. This is primarily due to the feedback which assessment provides to students regarding the nature of learning in general, and their own learning in particular (Rowntree, 1977, pp. 72-73). 'Feedback, or "knowledge of results", is the life-blood of learning' (Rowntree, 1977. p. 24).

That students respond to such feedback imposes significant conditions on the nature of assessment data and provides a platform for criticising the ' 'all-talking, all-singing, all-dancing' uni-dimensional grade' (Rowntree, 1977, p. 198) and for supporting profiles of achievement, which are essentially multivariate data, and non-numerical indicators of achievement such as verbal descriptors.

> Feedback is the principle mechanism through which assessment for formative purposes is realised. It refers to information returned to the student AFTER tasks have been performed, and is idiosyncratic, that is, feedback is referenced to the specific performance of a student. Feedback is not only an integral aspect of formative assessment but an essential element of the teaching-learning process. Further, it involves more than information, printed or otherwise, concerning the correctness or incorrectness of a response (Findlay, 1987, p. 4).

The impact of assessment on learning is well acknowledged.

> Assessment procedures have a profound capacity to direct and shape student learning. They send powerful messages to students and teachers about the kinds of courses and learning that are valued and considered worthwhile. The potential of assessment

procedures to direct learning in the senior secondary school has been an underlying theme in all recent reviews of post compulsory education. (Masters & Hill, 1988, p. 274).

The relationship between assessment and learning can be understood in terms of the operation of the classroom as a dynamic system. Biggs and Moore (1993, pp. 447-453) propose the 3P model of teaching and learning which serves to outline the major components of the classroom system which influence learning, and further to identify the characteristics of these components which contribute to good learning (primarily construed as deep, qualitative learning (Biggs & Moore, 1993, pp. 20-26). According to the 3P model the components of the classroom system are threefold: presage, process and product (hence 3P). Presage components which are 'those aspects of the teaching and learning contexts that existed prior to the immediate action in the classroom'(Biggs & Moore ,1993, p. 449). Presage factors are categorised as either 'student characteristics' or pertaining to the 'teaching context'. Assessment is a presage factor in a number of different ways. A student's previous experience with assessment will influence that students conception of learning, and how that student will approach the task of learning. Also, the students expectations of success and failure, another presage factor, is shaped by the students conception of the assessment system by which that student will be judged. Also, teachers, particularly if judged themselves by the student results on assessment tasks as in the case of high stakes testing situations (Smith, 1991), will teach according to the requirements of the assessment system irrespective of the implications for good learning.

Process components account for the way students approach and engage a particular learning situation within the classroom. Students adopt strategies and develop motives in accordance with their perception of the nature and requirements of a learning situation, and student perception is shaped by their conception of learning, which is inturn shaped by their previous experience and future expectation of assessment.

Product components refer to student achievement defined by the assessment system. The outcome of learning formally denoted by the assessment system, defines the very purpose of the classroom by defining 'success' for a set of learning episodes. This provides feedback, shaping conceptions and approaches to learning which become manifest in future learning situations.

Presage, process and product components (hence 3P) are interrelated and form a dynamic system which respond to assessment.

Assessment provides feedback, and feedback is an important issue for understanding learning. This is because learning occurs via feedback, and the notion of feedback locates assessment within a broad frame of reference, i.e., facilitating learning. Feedback occurs at two levels. The first is the classroom level. Classroom environments are components of larger systems (Biggs, 1991, p. 227), and respond to external influences. They are essentially open systems (Doll, 1989, pp. 245-246) and being open systems they evolve to accommodate elements of their environment. The impact of high stakes testing on teaching and learning, for example (Meisels, 1989; Smith, 1991), drives the classroom system towards teaching to the test, at the expense of other educationally worthwhile goals (Shepard, 1989). High stakes testing becomes a presage component of the classroom system, particularly when teachers are evaluated by the scores which their students attain on the tests, inturn impacting processes and products within the classroom:

> test scores, rather than playing the role of passive indicator variables for the state of the system, become the currency of feedback within an adapting educational system. The system adjusts its curricular and instructional practices, the students adjust their learning strategies and goals, to maximise the scores on the tests used to evaluate the educational outcomes, and this is particularly true when the states are high (Frederiksen & Collins, 1989, p. 27).

The second level of feedback is with the students themselves. Students are systems existing as nested components within the classroom system. Students respond to feedback and cues within the classroom, and the 3P model is most useful when accounting for elements which provide desirable feedback. Desirable feedback is that which promotes desirable learning processes, and desirable processes, according to Biggs and Moore (1993, p. 313), are facilitated by a deep approach to learning.

The nature of 'desirable processes' requires further elaboration. The classroom exists to foster student learning, and desirable processes of learning are to be understood in terms of the nature of learning itself. Traditionally there are two approaches to understanding learning: one emanating from research in cognitive psychology and another from research in education (Richardson, 1987, p. 3). These two approaches seem to be complimentary (Richardson, 1987,

p. 6). Cognitive psychology has sought to develop general theories regarding the nature of learning, whereas educational research has tended to focus on how students conceive learning and how they approach learning situations.

Within the field of education, the nature of learning has been studied from a phenomenological perspective, also called a second order perspective (Marton, 1981), and has sought to understand the phenomenon of learning from the perspective of the learner in terms of the way the learner perceives and approaches the task of learning (Marton & Saljo, 1976, 1984). This approach has endeavoured to uncover the attributes of successful students, and how they function within educational institutions.

Three important frameworks for describing student learning which have emerged from educational research are the stages of intellectual development proposed by Perry (1970), the conceptions of learning (Marton & Saljo, 1984), and the approaches to learning (Marton & Saljo, 1984). Taken together, these frameworks present the view that learning is primarily determined by the way the student perceives the task of learning, which influences how the student approaches learning, and in turn determines the nature of the learning outcomes, i.e., how the student functions within a task environment. The relationship between learning conception, approaches to learning and learning outcome is supported by empirical data (Van Rossum & Schenk, 1984).

One weakness of educational research, however, is that although:

> we are offered what are extremely sophisticated taxonomies of different possible approaches to learning...educational research has been fairly weak in providing well articulated general theories of human learning. The taxonomies are descriptive rather that explanatory when it comes to the nature of learning itself, and we are told very little about what actually happens when a student learns (Richardson, 1987, p. 5).

The nature of good learning is essentially a process of development which is manifest in the ability of an individual to function within a task environment. Understanding the interrelationship between what a person knows and what a person is capable of doing is at the heart of understanding good learning. Learning is the process of developing a well structured knowledge base, and 'Knowledge powerfully affects, and is reflected in, behaviour' (Ormell, 1992, p. 27), i.e., functioning in a task environment. Furthermore, it is the structure of one's knowledge, the interrelationship between concepts, and the extent and dynamism of ones cognitive schema which determine expertise (Larkin & Reif, 1979; Anderson, 1982), the advanced stage of learning.

> As knowledge develops, its structure eventually goes beyond classification of concepts and beyond networks of first order relations to include complex systems of multiple relationships as well as organising frameworks for interpretation and action, which are called schemas. Schemas, as the name implies, are schematic mappings of relationships in any one of a number of forms, such as propositional, iconic, or kinaesthetic (Messick, 1984, p. 218).

The concept of the nature of learning as being the development of a well structured knowledge base (a complex system of concepts), has arisen within the field of cognitive psychology and has been applied to other fields in cognitive science. The focus on the knowledge base as the origin of expert performance, for example, has been particularly fruitful for the development of Artificial Intelligence expert systems (Taylor, 1991, pp. 164-165).

Another explanatory framework which locates feedback as a central component of the development of a well structured knowledge base is Pask's conversation theory (Pask, 1976, 1984). Entwistle (quoted in Holmberg, 1989, p. 49) provides the following comment:

> Essentially this theory describes learning in terms of a conversation between two representations of knowledge. In the most familiar situation these representations reflect the cognitive structures of two people, the teacher (or subject matter expert) and the student. Learning takes place through a dialogue between the two and, in conversation theory, understanding has to be demonstrated by applying that knowledge to an unfamiliar situation in a concrete non-verbal way...Reproductive responses based on memory are not accepted as evidence of learning.
>
> Learning need not, however, involve an interaction between the cognitive structures of two people. The student may converse silently with himself in trying to understand a topic, or he may interact with a formal representation of the knowledge structure and supplementary learning materials which have been specially designed to facilitate understanding of the chosen subject-matter area.

Learning is influenced by 'conversation' with the assessment procedures and the assessment data generated by such procedures. The assessment data must communicate meaningful information to the student regarding the nature of learning and their developing cognitive structures.

The literature reviewed for this section locates assessment as a component of a dynamic system which responds to feedback, and as a component which is powerful enough to direct the shape of classroom learning, for better or worse. The classroom functions as a system and because knowledge is constructed by the learner, the students perception of the learning environment, infused with the curriculum agenda manifest in the assessment paradigm, provides the student with knowledge about the nature of the learning situation; the student constructs a conception of learning from this knowledge which inturn influences the students approach to learning. The process of learning is determined by the way the student approaches learning, and the development of the student's knowledge base is determined by the approach the student adopts (Van Rossum & Schenk, 1984). The structure of the students knowledge base is manifest in the learning outcomes evident through the way the student functions in the task environment (Biggs & Moore, 1993; Frederiksen, 1984, pp. 367-370).

2.2.2 The macro-functions of assessment

Macro-functions of assessment are primarily concerned with selection and decision making processes made for and by people and bodies outside the classroom. The Scott Report (Scott et al., 1978, p. 40) identifies two macro-functions of assessment:

- to assist students and parents in determining future education and employment pathways;
- to provide, for other educational institutions and employers, an indication of the suitability and readiness of the student to undertake further education and/or employment.

Docking (1987, p. 10) identifies the following five macro-functions of assessment:

- System evaluation (accountability);
- Student advancement (selection and promotion);
- Student certification (final);
- Student reporting (periodical);
- Teacher evaluation (curriculum and teaching).

Rowntree (1977) also catalogues two:

- selection of candidates for various types of educational opportunity or career (p. 16);
- maintaining standards (p. 20).

Macro-functions of assessment facilitate selection and certification processes which ultimately require an ordering of candidates. The nature of these macro-dimension processes implies that the assessment data generated be able to be ordered, and ordered on a variety of different univariate measurement scales.

Discussions of the features of the macro-functions of assessment, such as the issue tertiary entrance and the nature of certification practices, can proceed only so far without being grounded in the assessment procedures of a particular education system. In Australia, the administration of education is a state and territory issue, so that the largest unit of analysis of the detail of micro-functions is the state and territory education systems. The detail of the micro-functions of assessment differs from state to state (BSSSS, 1991), as one would expect if systems are allowed to evolve independently. Consequently, macro-function literature written outside the specific concerns of the macro-functions of assessment for Queensland education (Creswell, 1987; Docking, 1987; Cooksey, 1993) is of limited value to this study since discussions of the implications of macro-functions on the data processing of MiFAD depend to a large extent on the particular nature of the MiFAD and MaFAD defined within each education system.

The major elements of the assessment procedures in Queensland senior secondary school education have been outlined in chapter one. These elements provide the macro-functions of assessment within Queensland being considered in this study, namely, certification (R6) and tertiary entrance (SAIs, OPs and FPs). Specific detail of the structure and issues surrounding the macro-functions of assessment is provided in various documents and reports produced by the Queensland assessment authorities (BSSSS, 1993b; TEPA, 1993a, 1993b) but are not relevant the specific focus of this study.

2.3 Criteria-based assessment

Criteria-based assessment is an assessment paradigm which was introduced into the Queensland secondary school assessment system following recommendations from the Scott Report (Scott et al., 1978). It was hoped to be a solution for certain problems which were identified within the Queensland school-based assessment system (Campbell, 1976; Fairbairn, McBryde, & Rigby, 1976). These problems had emerged in the wake of the structural changes which took place following the recommendation from The Radford Report (1970) that public examinations be replaced by school-based assessment. (For an overview of the changes to the assessment system in Queensland secondary schools between 1964 and 1983 see Clarke (1987).)

ROSBA presented a synthesis of the criticisms found by Fairbairn et al. (1976) and Campbell (1976), listing 28 criticisms in all (Scott et al., 1978, pp. 5-7). Among these appeared the following eight (the numbers are the order as they appear in the Scott Report):

13 Tests and examinations remain the imperative of school life.

14 Assessment is almost exclusively concerned with the recall of academic knowledge.

15 Low priority has been given ... [to] feedback to amend teaching strategies and to diagnose student weaknesses.

16 Testing and ranking of students have increased in frequency and are having a detrimental effect on students, teachers and school administrators.

18 Continuous assessment together with relativistic ratings have generated anxiety and hostility in students.

20 The promise of freedom in evaluation practices remains largely unfulfilled.

21 Students believe the distribution of ratings to their school are pre-determined and this has led to a decline in teacher-student relationships.

27 A marks 'fetish' has developed leading to unhealthy competition.

The assessment paradigm which was operating within the school-based assessment system at the time of ROSBA was norm-based assessment. Norm-based assessment (i.e. norm-referenced assessment) is a statistical method of interpreting test scores which involves scaling student achievement to norm means and standard deviations. Students are rank ordered onto an interval scale continuum and grades are allocated according to where the students are ranked relative to the other students in the cohort. (The theoretical and technical features of norm-referenced assessment are to be found in textbooks on educational and psychological measurement, e.g., Anastasi, 1988).

Within a norm-based assessment paradigm grades are essentially pre-determined because they are defined with respect to pre-defined areas under the normal curve (criticism 21), the assessment process is primarily rank ordering (criticisms 16 and 18), facilitated by univariate scores (criticism 27). The numerical data required to facilitate norm referencing restricts what can be assessed (criticism 14), how it can be assessed (criticism 20), the frequency of assessment (criticism 13) and the quality of the feedback to the teacher and the student (criticism 15). The problems with the assessment system were placed squarely at the feet of the norm-based assessment paradigm (Scott et al., 1978, p. 36) and consequently it was replaced by competency-based assessment (p. 36) which later evolved into criteria-based assessment.

Educators are aware of the limitations of norm-referenced measurement procedures. Norm-referenced measurements provide little information about actual standards achieved (Power, 1986, p. 267), and make it difficult to determine long term drifts in achievement due to the relative nature of the grades (Sadler, 1987, p. 192). Also, grading systems which are based on a predetermined distribution may be artificially creating a shortage of high grades (Deutsch, 1979) thereby creating an undesirable competitive environment within the classroom.

The influence of norm-based assessment on education systems is extensive:

> Most of the destructive effects arise from the vagueness and relativity of systems that determine standards by reference to norms rather than by reference to competencies, and include student alienation, stress and anxiety, excessive competitiveness, extrinsic motivation, curriculum and assessment invalidity, meaningless grades, loss of accountability, restricted individualisation, ambiguity in goals and standards, test 'gamesmanship', narrowed curriculum and restricted expectations (Docking, 1987, p. 10).

The use of the normal curve for grading students is extensively criticised elsewhere:

> How did teachers ever suppose that the statistician's bell-shaped curve ought somehow to be imposed on the results of their work? The famous curve was developed to describe the distribution of natural phenomena...The bell-shaped curve is descriptive of raw or unselected phenomena, and then only if vast numbers of cases are used. When the teacher receives his pupils, their distribution with respect to some characteristics may follow the "normal curve". But having received his charges, the skilful teacher sets out as fast as he can to destroy the natural state of affairs (Bresee, 1976, p. 108).

It is not surprising that norm-based assessment became problematic after the transition from public examinations to school-based assessment. Under school-based assessment the responsibility for designing, implementing and interpreting educational assessment rests ultimately with the teachers in the school. Norm-based assessment is an assessment paradigm which essentially serves 'large-scale assessment purposes' (Stiggins & Bridgeford, 1985), which are characterised by the requirement to rank order students based on their achievement in the assessment component of the curriculum. Such rank ordering is used, among other things, to facilitate selection processes for entrance into tertiary education institutions. Inside the classroom, assessment must serve an entirely different set of purposes, namely, the facilitation of teaching and learning (section 2.2). Here assessment must serve 'teachers' primary measurement needs' (Stiggins & Bridgeford, 1985) and the application of large-scale assessment models such as norm-based assessment creates complication and confusion within the classroom (Imre, 1983).

The study of criteria-based assessment poses some significant challenges. Although it is 'located firmly within the criterion-referenced tradition ... [it] is sufficiently distinct from the most fully developed existing varieties in the USA for it to require independent developmental work' (Sadler, 1986a, p. 4). There are textbooks about the technical features of criterion-referenced testing but no such textbook exists for criteria-based assessment. The scrutiny of criteria-based assessment is hampered by its developmental origins and that the process of its early development, which laid the foundations for the assessment paradigm of today, occurred 'in the field' and largely outside discussion within the educational literature. A thorough study of criteria-based assessment, the kind required to facilitate independent developmental work, would involve a detailed documentary analysis of Board literature, subject syllabi and school work programs, together with research into the varieties of current practice. A comprehensive work of this nature does not yet exist.

The confusion surrounding the nature of criteria-based assessment has not gone unnoticed:

> It is uncertain where the notion of criteria-based assessment came from. It seems to be a weak version of criterion or domain-referenced assessment...However 'criterion-based [criteria-based] assessment' is not a term used in the measurement literature. It can only be assumed that it is an example of bureaucratic hyper-rationalisation. What is to be done

> is defined in grand and vague enough terms for us all to agree to act, but not to know what to do (Power, 1986, p. 269).

An example of the lack of rigour within the development of criteria-based assessment surrounds the issue of terminology. The Scott Report refers to 'competency-based assessment' although its reference to objectives and mastery (Scott et al., 1978, p. 36) suggest very strong links with criterion-referenced testing. The report does not cite from the educational literature, nor does it reference existing examples of competency-based assessment. The term competency-based soon disappeared and was replaced with two terms, 'criteria-based' and 'standards-based' (also called 'standards-referenced'). Sadler (1986a, p. 6) initially argues for the use of the term standards-based (it also seems that the notion of 'competency' used in the Scott Report means the same thing as standard, e.g., 'levels of competency represent standards of performance' (Sadler, 1986a, p. 6)):

> because standards presuppose criteria, an assessment system based on standards incorporates necessarily the concept of criterion-referencing. For this reason the new scheme could be referred to generically as "standards-based assessment".

He later uses the term 'criteria-based' (Sadler, 1986b, p. 9), and finally settles for standards-referenced (Sadler, 1987). It is worth noting that although standards-based and standards-referenced are synonymous, criteria-based and criterion-referenced are not.

Another confusing issue arises with the terms 'criteria' and 'standards' which are the two fundamental concepts within criteria-based assessment. Sadler (1987, p. 194) provides the following definitions:

> criterion - A distinguishing property or characteristic of any thing, by which its quality can be judged or estimated, or by which a decision or classification may be made (From Greek kriterion, a means for judging).
>
> standard - A definite level of excellence of attainment, or a definite degree of any quality viewed as a prescribed object of endeavour or as the recognised measure of what is adequate for some purpose, so established by authority, custom, or consensus (From Roman estendre, to extend).

These are the meanings attributed to the terms within criteria-based assessment, but these meanings are not held universally:

> This distinction between criteria and standards is not made universally, but is to be found in some of the literature on educational assessment. In fact, Glass (1978) traced a shift in the meaning of the term criterion and it has been used in educational measurement. Originally, a criterion referred to a characteristic or dimension of performance, but since the time of Glaser (1963), a criterion has been used to mean the particular score that is taken to designate competence or mastery. The 'criterion' in criterion-referenced testing is invariably a numerical cut-off...[which] would be better called a standard (Sadler, 1987, p. 195).

Despite the ever evolving nature of criteria-based assessment, that it has evolved largely within schools from syllabus guidelines and not within the educational literature, and that the assessment paradigm differs from subject to subject, there is enough literature to identify and clarify some of its generic features. Perhaps the most important documentation is a series of 21 discussion papers produced by the Assessment Unit of the Board during 1986 and 1988, which contain the essential elements of criteria-based assessment.

The essential feature of criteria-based assessment is that the assessment data which is generated within this assessment paradigm is multidimensional and not necessarily numerical. It is multidimensional because achievement is referenced to a set of criteria, which are essentially a set of domains. Overall achievement is a profile of achievement which consists of achievement within each of the individual criterion. It is not necessarily numerical because achievement in some of the criteria does not lend itself to measurement using an interval scale.

These two features of criteria-based assessment, namely, multidimensionality and freedom from the strict use of measurement scales, are both its strength and its weakness. Its strengths lie in the nature of the information which it provides to teachers and students about student performance. Its weaknesses lie in the technical difficulties associated with rank ordering multidimensional data.

Criteria

The original criteria set out in the Scott Report (Scott et al., 1978, p. 30) to form the basis of achievement within criteria-based assessment were four categories of objectives:

- process objectives (the cognitive skills to be developed in relation to the discipline);
- content objectives (relating to the body of factual knowledge to be acquired);
- skill objectives (the practical skills required by the discipline);
- affective objectives (the attitudes, values and feelings, the discipline aims to develop which may be manifest in overt performance).

These categories of objectives form domains, the elements of which are specific objectives. The domains in criteria-based assessment, however, are quite different to the domains in criterion-referenced testing. Under criterion-referenced testing the domains are sets of tasks which are generated from behavioural objectives. The tasks used for the assessment of an objective are a sample of an hypothetical population of tasks generated from the objective. Student achievement is referenced to the number of tasks which they perform correctly within each domain sample, an indicator of how well the student would perform in the population of tasks for that objective. Mastery is defined by a certain cutoff number of tasks answered correctly.

The domains in criteria-based assessment are less prescriptive and the objectives contained therein are not intended to be behavioural objectives. The domains contain statements, couched as objectives, which refer to a class of tasks or activities. Tying assessment to behavioural objectives restricts the assessment to low level cognitive outcomes, inturn restricting, via backwash, the teaching and learning to low level cognitive activities (Power, 1986, pp. 267-269).

The separation of performance into a set of criteria domains seems to be a necessary condition for an assessment system which seeks to reference student achievement to a task set rather than relative to the performance of other students. For the assessment to be valid and the measurements to be meaningful, the assessment instruments must be measuring an homogeneous entity: the instruments must be measuring the same thing. Criteria are intended to represent domains of homogeneous 'outcomes', and achievement within a subject is essentially multidimensional. Hence multiple criteria for a subject.

Cresswell (1987) analyses the concept of a domain as it has been defined by the Secondary Examinations Council (SEC) in the United Kingdom. The SEC, responsible for the introduction of grade criteria into GCSE examinations, set up Grade Criteria Working Parties whose task it was to divide subjects into domains. According to the SEC, a domain is 'a collection of the elements of a subject that forms some reasonably subset of the skills and competencies needed in the subject' (SEC, 1984).

The reason for the introduction of domains was because of the 'perceived difficulty of establishing sufficiently explicit criteria in a subject as a whole' (Cresswell, 1987, p. 247). Subjects are complex and competency is multidimensional. This is an example of the need for multiple criteria (domains) within a subject once educational achievement is no longer referenced to statistical norms. The Secondary Education Council (cited in Creswell, 1987) put it this way:

> It is unlikely that any working party will be able to produce GRC [Grade Related Criteria] that will be sufficiently explicit as to the knowledge and skills achieved by a candidate at a particular grade in the subject as a whole. The GRC will, therefore, need to specify performance in terms of achievement in prescribed domains (p. 247).

Assessment systems which reference achievement within a subject to multiple criteria seek to overcome the ambiguity of the overall grade. The ambiguity arises because of the unidimensional nature of grades. Competence is not a unidimensional entity, therefore there must exist a set of different 'profiles' of competency levels across each of the subject domains for at least one of the grade categories. This means that at least one grade category can be achieved in more that one way, hence ambiguity.

Standards

Criteria-based assessment represents student achievement in terms of a set of attained standards across a set of criteria, one standard awarded per criterion. Within a single criteria there exists a set of ordered standards. Achievement within a criterion can be conceptualised as a position along a continuum of achievement, and a standard can be interpreted as a category represented as an interval along the continuum.

The simplest method of defining a set of standards is to use a set of cutoffs along a numerical scale (Sadler, 1987, p. 198). The continuum of achievement is represented by the numerical scale, and a standard, say si, is defined as an interval, i.e., [x_{i-1} , x_i]. If a student receives a score of x such that $x_{i-1} \leq x \leq x_i$, then the student is awarded a standard of si. This is the approach adopted in both criterion-referenced testing and norm-referenced testing, the difference being the method of defining, or 'referencing' the interval [x_{i-1}, x_i]. This approach is facilitated by the use of measurement scales to represent the continuum of achievement.

Perhaps one of the most significant differences between criteria-based assessment and both norm- and criterion-referenced testing is the departure from the use of measurement scales in assessment. The use of measurement scales for educational assessment has been scrutinised (Docking, 1987; Sadler, 1988) and some concerns have been raised.

> An analysis of the underlying assumptions shows that numerical marking systems enjoy a status that is higher than they strictly deserve. The use of marks in criteria-based assessment is inappropriate for two sets of reasons. Firstly, the assumptions [necessary for the use of measurement scales] are not generally satisfied in any form of school-based assessment, and secondly, the use of marks as currency in grade-exchange transactions diverts attention away from criteria, standards, and the process of qualitative appraisals, and to that extent is educationally counterproductive (Sadler, 1988, p. 10).

Alternative approaches to numerical cutoffs for promulgating standards are the use of exemplars (Sadler, 1987, p. 201) and verbal description (Sadler, 1987, p. 200). Irrespective of the method for determining standards, however, since there exists a finite set of standards within a given criteria, there is still the need for a finite set of codes, such as letter codes (A, B, C,...) or numbers (1, 2, 3...), or other symbols (S1, S2, S3,...) which essentially refer to standard categories. It is important to realise that 'having a set of standards does not necessarily imply an interval scale' (p. 197) so that the movement from an E a D is not necessarily the same 'jump' as the movement from a B to an A.

Levels of achievement

Despite the emphasis on criteria and standards within criteria-based assessment, student profiles of standards across a set of criteria must be summarised into one of five level of achievement bands: VHA (Very High Achievement), HA (High Achievement), SA (Sound Achievement), LA (Limited Achievement), VLA (Very Limited Achievement).

These level of achievement bands are overall grade categories representing overall achievement. Under norm-based assessment the overall grade categories were, in ascending order of achievement, integers from 1 to 7. The reasons for the decision to replace the seven norm-based numerical grades within five level of achievement grades were two fold:

> First, we wish to avoid any suggestion of equivalence between the assessment obtained through norm-based procedures and competency-based [criteria-based] procedures. Second, we believe that the competency scale should be divided into meaningful unambiguous categories. It is our contention that students can be ranked more reliably into five bands of competency than they can into seven. (Scott et al., 1978, p. 37)

Furthermore, the shift from the use of numbers to verbal descriptions was deliberate:

> Some invited reactors to our proposals rejected the proposed verbal descriptions of performance in favour of either a numerical or alphabetical category system. We reject these arguments on the grounds that a numerical system would lend itself to the invalid malpractice of summing levels of attainment, as currently occurs with semester points in senior secondary assessment, while an alphabetical system would require a translation step to a description in terms of competency - a requirement that we believe should be avoided (Scott et al., 1978, p. 37).

Irrespective of codes or the number of categories, all sets of overall grade categories are ordinal in nature. Despite this, two different types of overall grades can be identified: those which are defined as intervals, or regions, along an interval-scale continuum, and those which are distinct classes (Cresswell, 1987, p. 257). Cresswell actually identifies four versions of the overall grade, two versions per type. Version 1 and 4 are of the first type, version 2 and 3 of the second.

> Version 1
>
> (i) Each component of the examination is marked.
> (ii) The marks awarded on the separate components are added (after scaling if necessary) to obtain a total mark.
> (iii) The scale of total marks is partitioned into regions, each of which corresponds to a grade.
>
> Version 3
>
> The domains are first graded ... but the domain grades are not then combined arithmetically to produce aggregated grade scales, instead combination rules are determined directly by a consideration of the patterns of achievement which are judged to correspond to the different overall grades (Creswell, 1987, p. 257).

Levels of achievement are version 3 grades since they are defined according to decision tables which specify combinations of standards for each grade. Therefore, at no stage of the criteria-based assessment grading process is overall achievement mapped onto a unidimensional numerical scale. Hence, the issue of ordering CBAD on interval scales, as required for the SAIs, is not addressed by criteria-based assessment grading.

2.4 Ordering multivariate data

This study is investigating an issue which is located within the broader issue of generating ordered distributions of students using educational assessment data (section 1.1). In particular, this study is concerned with the nature of ordering mechanisms for ordering multivariate data. Consequently, the nature of order, and the various approaches taken to formalise the concept of order, has been reviewed for this study.

The concept of order has been researched across four fields: philosophy (section 2.4.1), set theory (section 2.4.2), statistics (section 2.4.3), and utility theory (section 2.4.4). Of these, utility theory, or more specifically, the topic of multiattribute utility functions located as an extension of utility theory, has provided the direction for this research. Some of the important findings from this literature will now be presented.

2.4.1 Order in philosophy

Order is a central concept in Western philosophy, but despite this, Lorand (1992, p. 579) remarks that the concept of order itself seems to have received little attention in philosophical discussion. That the concept of order deserves scrutiny lies in the assumptions which empiricist and rationalist theories make about the nature of order. Such assumptions lie at the heart of these theories, and although there are disputes about the nature of order, disputes about 'theories of order are mainly about the kind of order found in nature, its status and origin, but not about the concept of order itself' (Lorand, 1992, p. 579). Hence, Lorand (1992) addresses the concept of order through a critique of the work done by Bergson.

It has been suggested (Bergson, cited in Lorand, 1992, p. 580) that order is neither a wholly subjective nor a wholly objective phenomenon but rather 'a certain agreement between subject and object. It is the mind finding itself again in things' (Bergsen, cited in Lorand, 1992, p. 580). 'Order reflects the way our mind operates, but the fact that this reflection is made possible carries information about the object itself' (Lorand, 1992, p. 580). Minds have choice and such choice is inherent, to varying degrees, in any ordering of objects.

Bergson's contribution to the debate about order lies in his argument against the concept of disorder (Lorand, 1992, p. 580). Traditional Western thinking claims there to be only one type of order and order is a triumph over disorder (implying the existence of disorder). Bergson, on the other hand suggests that there are really two different types of order and that disorder does not exist. These two types of order are called 'geometrical order' and 'vital order'. Both of these types of order emerge both from characteristics of the mind of the subject which is doing the ordering and characteristics of the object which is being ordered.

The details of Bergson's argument and Lorand's point of view are beyond the task of this section. Much of the discussion is not directly applicable to this study. However, the concept of geometrical order as the operation of the mind's intellect on unorganised bodies, and vital order as the operation of the mind's intuition on organised bodies, may help to explain the difference between objective scoring procedures and subjective marking processes in educational assessment.

According to Bergson, the first type of order, geometrical order, is the traditional type of order implicit in philosophy and science. It arises from the action of the intellect on unorganised bodies.

> Bergson offers many synonyms for geometrical order, each expressing a different aspect of the same line of thinking: physical order, automatic order, spatial order, intellectual order, the order of unorganised bodies (Lorand, 1992, p. 582).

Geometrical order is determined by an ordering principle which is imposed on a set of unorganised bodies (unorganised bodies may be separated and treated as a set of discrete parts; there is no inherent systems quality to such a collection of parts). A particular ordering of objects is the operation of an ordering principle, chosen by the mind, on those objects. Order is necessarily determined by an ordering principle although the ordering principle itself may not be necessarily determined by anything.

> In an ordered set, the location of the elements is determined by the particular ordering principle, even if the principle itself is not necessarily determined by anything. For instance, one may choose to arrange books according to the alphabetical order of their titles, or their author's names, or by the number of their pages; nothing dictates or decrees the chosen principle. But the location of each book is necessarily determined by the chosen principle (Lorand, 1992, p. 581).

The ordering principle is external to the bodies being ordered at the time of ordering; it is imposed on the set of discrete bodies.

Consider a spelling test. The ordering of a student cohort using the number of words spelt correctly by each student is an example of geometrical ordering. The ordering principle is defined prior to ordering and is imposed on the set of words spelt by each child. The set of words is an unorganised body.

Vital order, on the other hand, is 'the intuitive, natural, and positive order, the willed and creative order' (Lorand, 1992, p. 582). Whereas geometrical order reflects the operation of the intellect via the separation of entities into discrete parts, vital order reflects the operation of the intuition on wholes. Vital order is the order of organised bodies, where such bodies are systems of parts for which the separation of the parts would change the entity being ordered. The ordering principle for organised bodies is intrinsic to such bodies and emerges as the body is being ordered, not prior to and imposed on as for geometrical order.

The assessment of a student essay, for example, is perhaps an example of vital ordering. The essay is a complex whole, an organised body, and the ordering principle evolves as the body is being ordered. The degree to which the essay 'hangs together', to which it is consistent and coherent, is its degree of vital order.

This philosophical discussion on the concept of order, briefly reviewed here, may have implications for educational assessment, but the pursuit of these implications lies beyond the immediate focus of study. Assessment which reflects the quantitative accretion of bits of knowledge, and inturn orders quantitatively, seems to involve what Bergson refers to as geometrical order. Qualitative assessment seems to be aligned with the concept of vital order. If learning involves the development of a well structured knowledge base, and if assessment is instrumental in the process of learning, perhaps the call for qualitative assessment in fostering a qualitative concept of learning in students (Biggs & Moore, 1993, pp. 20-26) could be considered in terms of the need for vital order rather that geometrical order as the basis of educational assessment.

2.4.2 Set theory

Set theory is the mathematical/philosophical study of sets, where a set is 'a collection into a whole, of definite, well- distinguished objects...of our perception or of our thought' (G. Cantor, cited in Kamke, 1950, p. 1). CBAD forms a set; the objects in this set are sets of standards and codes descriptive of achievement. The concept of order is an essential idea in set theory, and the word 'order' is used adjectivally in the context of describing the properties of sets, i.e., 'ordered sets' and 'well-ordered sets'. Set theory was reviewed for this study in the hope that it might be able to address the issue of ordering data in multidimensional space.

An ordered set is defined in the following way:

> A set M is called an ordered set, provided that a relation, denoted by the symbol $\prec$, subsists between every pair of distinct elements a and b of the set, and only between distinct elements, and satisfies the following two conditions:
> 1. If a (not equal to) b, then either $a \prec b$ or $b \prec a$.
> 2. If $a \prec b$ and $b \prec c$, then invariably $a \prec c$; i.e., the relation is transitive (Kamke, 1950, p. 52).

Despite its rigour, or perhaps because of it, set theory was found to be of little value in providing a basis from which to analyse the issue of ordering multivariate data. It became apparent after some initial reading in the area that the theorems developed in set theory form a network of the fundamental properties of sets, and ordered sets, but did not directly address the issue of how to order points in multivariate space. Nevertheless, set theory did provide some insight into the nature of the problem of ordering multivariate data: it seems that such data does not form an ordered set, and therefore is not inherently ordered.

For example, in multidimensional space, points which may not equal to each other (i.e. not the same point), e.g., (50, 70) and (70, 50), may nevertheless not be related by $\prec$. However, this is the case in ordering multivariate data, and the question arises: do such points form an ordered set? According to set theory they do not because they do not satisfy the first of the two conditions in the definition. This suggests that sets consisting of such points are not ordered sets. In other words, multidimensional assessment data are not ordered because such data do not form an ordered set. Despite this, another theorem, called the well-ordering theorem, states that every set can be well ordered (Kamke, 1950, p. 112), implying that every set can therefore be ordered. This suggests that order can be introduced into multidimensional space, but the details of how this is done were not found in the set theory literature.

2.4.3 Statistics

Attempts to generalise univariate statistical concepts like mean, median, range and extremes to multivariate data encounter difficulties which are associated with the lack of order in such data:

> the lack of any obvious and unambiguous means of fully ordering, or ranking, observations in a multivariate sample appears as an obstacle to the development of statistical method: in particular to the extension to higher dimensions of areas of application, methodological advantages or general properties of univariate order statistics (Barnett, 1976, p. 318).

Despite the lack of any unambiguous ordering of multivariate data, there do exist various types of sub-ordering principles for multivariate data. Four types of sub-ordering principles have been identified (Barnett, 1976, pp. 322-327): marginal ordering (M-ordering), reduced (aggregate) ordering (R-ordering), partial ordering (P-ordering) and conditional (sequential) ordering (C-ordering).

Of the four sub-ordering principles, R-ordering seems to be the most relevant to this study. The other three sub-ordering principles addressed specific statistical issues which are not relevant to this research.

R-ordering is described by Barnett (1976, p. 322) in the following way:

> With this type of ordering each multivariate observation is reduced to a single value by means of some combination of the component sample values. The metric employed is frequently of the "generalised distance" type...the aim is to effect some restricted overall ordering of the multivariate sample.

The condensation of the multivariate data to a single value orders the multivariate sample through the ordering of the univariate single values which correspond to each multivariate variable. The multivariate data are mapped onto a univariate scale, hence order is imposed on the multivariate sample space. Different orders are obtained from different methods of condensation.

The investigation of R-ordering as a topic containing the seeds of a mathematical system of the type sought after in the research question (section 1.5) was not pursued. It became apparent after some initial reading of the literature addressing this topic (Wilks, 1963; Chernoff, 1973) that although the issue of ordering multivariate data is of importance to statisticians, much of the associated methodology addresses specific statistical issues. The specific nature of a mathematical system which orders multivariate data and which models preferences was not found in this review of the statistical literature.

2.4.4 Value functions

The concept of a value function arises in the development of multiattribute utility functions (Keeny & Raffia, 1976). Multiattribute utility functions are used to model preferences and tradeoffs in decisions using multiple attributes. These functions are associated with the following fields: utility theory (Berger, 1985), decision theory (Keeny & Raffia, 1976), and operations research (Thompson, 1967) where questions regarding the optimisation of outcomes based on multiple inputs are pursued.

There is reason to believe that these fields may be able to contribute to understanding and modelling the decision processes used in assessment systems for Board subjects in Queensland schools. This is particularly the case for multiattribute utility theory, since teachers are required to make decisions of overall value and to couch these decisions numerically, using multiattribute data. Unfortunately, many of the important concepts and much of the work in these fields could not be pursued in a limited study of this nature. Nevertheless, certain concepts have been borrowed from these fields and applied to the research question being addressed here. These concepts will now be explained.

To begin with, multiattribute utility theory is associated with utility theory. Utility theory is described by Berger (1985, p. 47):

> To work mathematically with ideas of 'value', it will be necessary to assign numbers indicating how much something is valued. Such numbers are called *utilities*, and *utility theory* deals with the development of such numbers.

One of the advantages with using numbers to represent value is the inherent ordering of the numbers: numbers (real or rational) form an ordered set and questions such as 'which object is more valuable?' can be addressed through the magnitude of the utilities assigned to each object.

In situations where objects possess multiple attributes and where decisions about the overall value of an object must be made, the multiattribute value problem arises (Keeny & Raffia, 1976, p. 66). The multivalue attribute value problem is about tradeoffs, where 'the decision maker is faced with the problem of trading off the achievement of one objective against another objective (Keeny & Raffia, 1976, p. 66). More specifically, a decision maker is faced with the problem of choosing how to act such that the consequences of this action are optimised. The act occurs in the 'act space' (A) and the consequences of the act occur in the 'consequence space'. Furthermore, the consequences are multidimensional, so that although the consequences of an act, a, may be known in the consequence space, i.e., X1(a), X2(a), ..., Xn(a) where Xi(a) is the consequence of a for the ith attribute, since different acts are projected into a multidimensional space, one cannot choose one act over another since the consequences are not ordered.

> Roughly, the decision maker's problem is to choose a in A so that he will be happiest with the payoff X1(a),..., Xn(a). Thus we need an index that combines X1(a), ..., Xn(a) into a scalar index of preferability or value. Alternatively stated, it is adequate to specify a scalar-valued function v defined on the consequence space with the property that $v(x_1, x_2, ..., x_n) \geq v(x_1', x_2', ..., x_n') \Leftrightarrow v(x_1, x_2, ..., x_n) \succeq (x_1', x_2', ..., x_n')$ where the symbol $\succeq$ reads 'preferred or indifferent to'. We refer to v as a value function. The same construct has many other names in the literature — ordinal utility function, preference function, worth function or utility function. Given v, the decision maker's problem is to choose a in A such that v is maximised (Keeny & Rafia, 1976, pp. 67-68).

The situation in educational assessment which is being addressed here is not one of deciding an act, a, in an act space but one of ordering 'outcomes' in an assessment space. In this way the outcome is analogous to the consequence space in the multiattribute utility problem. Value functions order multivariate consequences through defining equivalence in the consequence space. Equivalence is defined in the consequence space using 'indifference curves' (Keeny & Raffia, 1976, p. 79) which pass through all points in the space which are deemed to be equivalent.

> We imagine that through any point x in an n-dimensional consequence space there is an indifference surface connecting all points that are indifferent to [equivalent to] x. These indifferences will be curves for n = 2 (Keeny & Raffia, 1976, p. 79).

Using the concept of an indifference curve, value functions model preferences in a multidimensional space (Keeny & Raffia, 1976, pp. 80-81) (see research question part 2), and necessarily order all points in the multidimensional space (see research question part 1). This research explores the design and application of value functions to a multidimensional percentage-based assessment space, and the use of such functions to develop independent OS for percentage-based assessment systems. The research will be outlined more fully in chapter 3.

2.5 Summary

The literature reviewed for this study has been organised into three sections. The first section (section 2.2) addressed the micro- and macro-functions of assessment together with the nature of the impact which these different functions have on the type of data generated by the assessment system. Micro-functions of assessment (section 2.2.1) requires that assessment data provides feedback to teachers and students to facilitate the learning process. Such data is referenced to the task environment via criteria and standards which are communicated using a range of coding systems including non-numerical codes. Micro-function assessment data (MiFAD) is multivariate data. Macro-functions of assessment involve the ordering of students for certification and selection processes. These functions require MiFAD assessment data to be ordered. The second section (section 2.3) examined the elements of the criteria-based assessment paradigm and the third section (section 2.4) considered the issue of ordering multivariate data.

CHAPTER THREE
RESEARCH METHODOLOGY

3.1 Purpose of Research

The purpose of this research is to explore the issue of generating macro-function assessment data from micro-function assessment data. More specifically, this research considers the issue of ordering multivariate educational assessment data for percentage-based assessment systems, and in particular, the nature of the ordering mechanisms for ordering systems within such assessment systems.

This study is focused by the following research question:
RESEARCH QUESTION:

Is there a mathematical system which:

(1) orders all points in a multidimensional percentage-based assessment space,

(2) models preferences within a multidimensional percentage-based assessment space,

(3) can be used to construct an independent ordering system for a percentage-based assessment system?

The research undertaken here to address this question is not an empirical study, although empirical data are collected and used in one section of the research (section 4.5). This is because the research question, as framed for this study, does not invite empirical investigation. This question seeks a mathematical system which possesses certain properties, i.e., orders multivariate data, models preferences in ordering of multivariate data, and as such is concerned with the properties of an abstract entity, namely, a mathematical system. The nature of order can be considered through non-empirical speculation, and although the specific nature of teacher's preferences of profiles of achievement are an empirical issue, generic properties of a mathematical system which models preferences can be identified without an appeal to specific preferences and decision structures.

Nevertheless, this study emerges within an educational context and seeks to respond to that context. It is also in a sense 'pre-empirical'. The broader aim of this research is to model ordering systems, which is an empirical issue since ordering systems are elements of an existing, functioning education system. However, there are features of ordering systems which can be identified and which require study prior to any empirical study. A non-empirical study of this nature may provide some conceptual tools which may be used to conduct further research of the empirical entity.

A survey of commonly used educational research methods (Jaeger, 1988) did not reveal a method to accommodate this research question. Instead, the question considered seems to be addressed by 'computer simulation experiments' (Thompson, 1967; Naylor, Burdick, & Sasser, 1967), a form of enquiry which develops and explores mathematical models as a prior stage of modelling an existing entity. Two issues arise for discussion. The first (section 3.2) is the nature of educational research, of what qualifies as educational research, and the location of this study within educational research. The second issue (section 3.3) is the nature of modelling and simulation and how this method appropriately responds to the research question being addressed here.

An outline of the research design for this study is provided in section 3.4.

3.2 Educational research

Research is characterised by a 'family of methods which share the characteristic of *disciplined inquiry*' (Shulman, 1988, p. 4). Disciplined inquiry is inturn characterised by its adherence to a set of principles and procedures which govern how it occurs, and more importantly, how it itself can be studied and verified (Cronbach & Suppes, 1969). To call inquiry research it must be disciplined inquiry, and in many instances this means that it follows the procedure of an established method.

Methods of research evolve and are applied within disciplines. Disciplines are organised around a set of issues, problems and questions. Furthermore:

> What distinguishes disciplines from one another is the manner in which they formulate their questions, how they define the content of their domains and organise that content conceptually, and the principles of discovery and verification that constitute the ground rules for creating and testing knowledge in their fields. These principles are different in the different disciplines (Shulman, 1988, p. 5).

What characterises research within a discipline is the application of a method recognised by that discipline, and since methods evolve to address a certain set of issues, subsequently to a set of issues with which the discipline is concerned. It is suggested, however, that educational research is not confined to the strict adherence to a subset of methods, because education is not a discipline but rather:

> *a field of study*, a locus containing phenomena...problems...and processes, which themselves constitute the raw material for inquiries of many kinds. The perspectives and procedures of many disciplines can be brought to bear on the questions arising from and inherent in education as a field of study. As each of these disciplinary perspectives is brought to bear on the field of education, it brings with it its own set of concepts, methods, and procedures, often modifying them to fit the phenomena or problems of education (Shulman, 1988, p. 5).

Research methodologies which are commonly employed in educational research are often identified as being either qualitative or quantitative. Qualitative methods include historical methods, philosophical inquiry methods, ethnographic methods and case study methods; quantitative methods include experimental methods, quasi-experimental methods and correlational procedures (Jaeger, 1988).

This study, however, does not fit easily into either of these two categories. The development and exploration of the properties of a mathematical system, as outlined in the research question, is not an empirical study, and these categories of methods are in the main oriented toward empirical investigation. Philosophical inquiry methods, such as conceptual analysis (Scriven, 1988), can be identified as pre-empirical, and in this way congruent to the thrust of this study, but these methods do not address this particular issue.

Nevertheless, although this research calls for a method, a form of disciplined inquiry not widely used in educational research, it is clearly educational research because the issue which it is addressing, namely, the structure and nature of ordering systems within educational assessment systems, is a problem located within the field of education.

3.3 Modelling and simulation

As described by Naylor et al. (1967, p. 1315), ‘computer simulation experiments are in effect experiments with a mathematical model’. Two distinct stages of computer simulation can be identified: the development of a mathematical model and the exploration of the operation of that model using computer programs designed using the model (Thompson, 1967, pp. 131-132).

The modelling stage of computer simulation involves the creation of a mathematical model or system. This system is developed with the intention that it possesses an identified set of properties, usually of an empirical entity, e.g., economic systems (Naylor et al., 1967). In this study, the mathematical model to be developed represents an ordering system which is an element of school-based assessment systems in Queensland secondary schools.

The ‘exploration of the model’ stage, also called simulation, involves the development of a computer program to simulate the operation of the model on simulated data. Data are processed by the model and the properties of the model are discussed through an analysis of the way the computer program, and hence the model, operate on the data. A variety of methods can be used to analyse the behaviour of the model, including correlation, scatterplots, and statistical hypothesis test.

This method of research is directly applicable to this research question. The research question seeks a mathematical system which possesses a given set of properties (section 1.5). Consequently, the development of such a system (modelling), the observation of its properties using computer simulation, and the discussion of the observed properties in light of those which the system is intended to have, is the approach adopted in this study to address the research question.

An overview of the research is as follows. A mathematical system will be developed in light of the properties outlined in the research question. Models will be developed from this system, and the operation of these models will be observed. The extent to which the models exhibit the properties of the mathematical system sought after in the research question, is the response of this research to the research question. The design of this research will now be addressed.

3.4 Design of the research

The research for this study is divided into four sections. The first section (section 4.2) is the first of two modelling sections. It involves the creation of a mathematical system to order points in a multidimensional percentage-based assessment space. The mathematical system uses the concept of an equivalence curve (indifference curve, Keeny & Raffia, 1976) and a target curve to define order in the space.

The second section (section 4.3) is the second modelling section of the research. It involves the development of specific models using the mathematical system developed in section 4.2. This section develops seven models, called OMu, and from each model is derived a value function called a p-algorithm. The first two modelling sections of the research are shown in Figure 3.1. The p-algorithms, p1 to p7, are for a 2-d assessment space (i.e., two criterion).

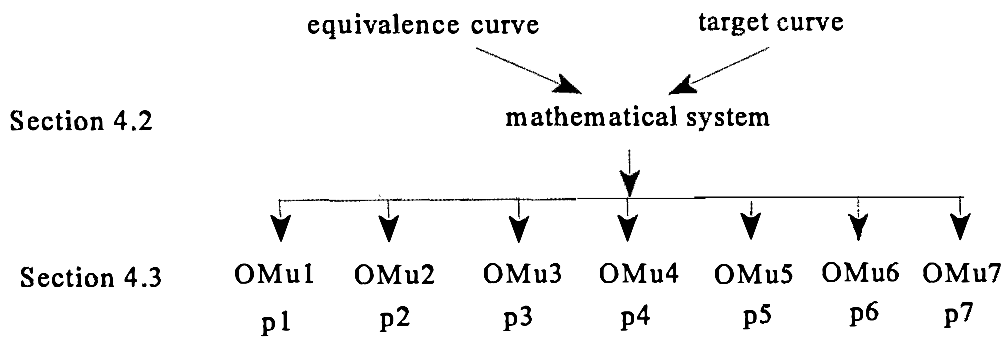


Figure 3.1: The modelling sections of the research

This models developed in this section are intended to:

(1) order all points in a multidimensional percentage-based assessment space, and

(2) model preferences within a multidimensional percentage-based assessment space.

In this way, the first two sections of the research together address the first two parts of the research question. The third section of the research explores the properties of the models in light of these requirements.

The third section (section 4.4) conducts six computer simulation investigations of the models (OMu) developed in section 4.3. The investigations can be divided into two sets. The first set consists of investigations 1, 2 and 3, and the second set of investigations 4, 5 and 6.

The first set of investigations use the seven p-algorithms developed in section 4.3 to map bivariate computer simulated data onto a univariate scale. In each investigation 100 data points are randomly generated by the computer, but the three investigations differ as to the regions of the assessment space in which the points are generated. For the first investigation the 100 points are generated throughout the entire assessment space. The second investigation generates the 100 data points in two regions of the assessment space, one called an extreme region and one called a non-extreme region. Extreme regions are regions of the assessment space which are at a distance from the target curve, whereas non-extreme regions lie around the target curve. The reason for this distribution of points is because the differences between the models developed in section 4.3 are manifest in the ordering of extreme points, and such differences reflect preferences (part 2 of the research question). This investigation illustrates the differences between the models. The third investigation generates data points only in a non-extreme region of the assessment space, illustrating the negligible differences between the mappings of the models for non-extreme regions.

The programs for the investigations are written in QBASIC and are contained in the appendices, together with the data for one run of each program. The appendices for each investigation are summarised in Table 3.1.

Table 3.1

Appendices for investigations 1, 2 and 3

Investigation	Program (Appendix)	Data (Appendix)
1	B	C
2	D	E
3	F	G

Each program does three things:

- generates the bivariate data,
- calculates a p-value (univariate) distribution for the data using each p-algorithm,
- calculates the correlation coefficient between each distinct pair of p-value distributions.

Two methods are used to investigate the similarities and differences between the ordered distributions generated by each p-algorithm for each investigation. The first is the correlation coefficient between all pairs of distributions. The correlation coefficient provides one measure of the degree of agreement between pairs of p-value distributions.

The second method is the use of scatterplots of pairs of p-value distributions. The points in the p-value scatterplots are the p-values from the two p-algorithms in the scatterplot calculated for each point in the assessment space. The distribution of points in the p-value scatterplots provides a visual representation of the similarities and differences between the p-algorithms, and in turn, the models from which they are derived. A line drawn across the p-value scatterplot, called an 'agreement line', indicates those points which are awarded the same, or similar, p-values by the p-algorithms in the scatterplot. The proximity of points to the agreement line represents the level of agreement. It is shown that points in the assessment space which lie on the target curve, fall on the agreement line for each p-value scatterplot.

The second set of investigations (4, 5 and 6) replicate investigations 1, 2 and 3 respectively, for 3-d versions of the 2-d p-algorithms. This set of investigations attempts to generalise the findings from the first set of investigations. Only four 3-d p-algorithms are used for these three investigations: p1, p3, p4 and p5. The appendices for each of these investigations is summarised in Table 3.2.

Table 3.2

Appendices for investigations 4, 5 and 6

Investigation	Program (Appendix)	Data (Appendix)
4	H	I
5	J	K
6	L	M

The fourth section (section 4.5) applies the four 3-d p-algorithms used in investigations 4, 5 and 6, to percentage-based CBAD obtained from the chemistry department of a Queensland secondary school for a cohort of students. In this way, this section addresses the third part of the research question since it constructs an independent ordering system for a percentage-based assessment system using the mathematical system and the models developed in section 4.2 and 4.3. The program used for this investigation is in appendix N and the data generated by the program is in appendix O. This section, however, does not represent a test of the mathematical system for representing actual decision structures used in ordering systems. It is merely an exploration of the possibility of conducting such an empirical study, and hence a possible avenue for further research.

CHAPTER FOUR
RESULTS AND DISCUSSION

4.1 Purpose of research

The purpose of this research is to explore the issue of generating macro-function assessment data from micro-function assessment data. More specifically, this research considers the issue of ordering multivariate educational assessment data for percentage-based assessment systems, and in particular, the nature of the ordering mechanisms for ordering systems within such assessment systems.

This study is focused by the following research question:

RESEARCH QUESTION:

Is there a mathematical system which:

(1) orders all points in a multidimensional percentage-based assessment space,

(2) models preferences within a multidimensional percentage-based assessment space,

(3) can be used to construct an independent ordering system for a percentage-based assessment system?

The results are divided into four sections. The first section (section 4.2) presents the development of a mathematical system for representing decision structures as ordering mechanisms within a 2-d percentage-based assessment system. The mathematical system developed here uses ideas from multiattribute utility theory (section 2.4.4), the main idea being that of the equivalence curve (section 4.2) which is essentially an indifference curve (Keeny & Raffia, 1976) renamed for the purposes of this study. This mathematical system will be shown (section 4.2) to yield models which order all the points in a 2-d percentage-based assessment space. Hence, this section will respond directly to the first part of the research question, namely: Is there a mathematical system which orders all points in a multidimensional assessment space? This section will attempt to develop a mathematical system which does this.

The second section (section 4.3) will develop a set of seven models, referred to as OMui (also OMu, or simply OM when it is obvious that the reference is not to either OMr6 or OMsai), where i = 1 to 7, which can be used to map percentage-based CBAD onto a univariate interval scale (U) for independent OS. A 'percentage algorithm' (also called p-algorithm) pi, where i = 1 to 7, will be deduced from each of the seven OMu for the 2-d percentage-based assessment space. The seven models developed here will provide a sample of models from the mathematical system developed in section 4.2. This sample will be used to ascertain differences in preferences between different regions of the assessment space for different models.

The third section (section 4.4) will explore the differences between the seven OMu developed in section 4.3 using computer simulated investigations. This section will endeavour to establish that different OMu represent fundamentally different preferences by exploring the differences in the ordered distributions for sets of randomly generated assessment data. This section consists of six computer simulation investigations and responds directly to the second part of the research question, namely: is there a mathematical system which models preference structures within a multidimensional percentage-based assessment space? This section will attempt to show that the models developed in section 4.3 from the mathematical system developed in section 4.2 exhibit different preferences for different regions in the assessment space.

The fourth section (section 4.5) involves the development of four independent OS using four of the OMu developed in section 4.3. It consists of a single computer simulation investigation, investigation 7, which applies 3-d p-algorithms for OMu1, OMu3, OMu4 and OMu5 derived from the 2-d p-algorithms for these OM developed in section 4.3. These independent OS will be used to generate R6 and SAI distributions for percentage-based assessment data for a cohort of senior chemistry students from a Queensland secondary school. This section seeks to show how the mathematical system developed in section 4.2 can be used to construct an independent ordering system for a percentage-based assessment system. This section will respond to the third and final part of the research question, namely: is there a mathematical system which can be used to construct an independent ordering system for a percentage-based assessment system? This section seeks to demonstrate the development of such ordering systems using the mathematical system developed in section 4.2 and the models derived from this system in section 4.3.

4.2 A mathematical system for generating ordering mechanisms

Let there be two sets, A and B, such that the elements of each set are real numbers (R) from 0 to 100.
i.e.

$$A = \{x : 0 \leq x \leq 100, x \in R\}$$
$$B = \{y : 0 \leq y \leq 100, y \in R\}$$

The combinatorial product A × B is the entire set of paired elements (x, y) which are formed by taking an element from A (x) and another from B (y) and placing the element from A before the element from B in the pair. The pairs constituting the combinatorial product will be called **points** and the elements within these points will be called **components**. Hence, the point (x, y) (in A × B) consists of the components x (in A) and y (in B).

The set A × B can be represented as the set of points contained in the plane formed by the two sets A and B along two axes. The plane is square and the elements of the combinatorial product are all of the points in the plane. Since the sets spanning the two axes of the plane are assessment measurements, this plane will be called the **2-dimensional assessment space** (2-d assessment space) or simply the assessment space when it is obvious that the situation is 2-d. This section considers only the 2-d case.The assessment space is shown in Figure 4.1.

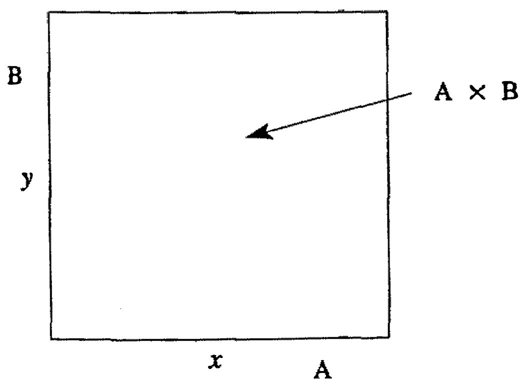


Figure 4.1: The 2-d assessment space formed from the sets A and B.

A unidimensional set is called an ordered set if and only if an ordering relation, denoted by the symbol $\prec$, subsists between every pair of distinct elements p and q of the set, and only between distinct elements, such that the following two conditions are satisfied:

1 If $p \neq q$ then either $p \prec q$ or $q \prec p$

2 If $p \prec q$ and $q \prec r$ then $p \prec r$

For the set of real numbers, or subsets thereof, the ordering relation is defined by the magnitude property of the natural numbers. The ordering relation derived from the magnitude property of the natural numbers is symbolised $<$. Since A and B are subsets of R, both A and B are ordered sets (ordered sets: section 2.4.2).

Ordering the set A × B

The assessment space is a set of multidimensional (bivariate) points consisting of real components. The task at hand is to develop a mathematical system to provide a basis for ordering the assessment space.

Since $x = 0$ and $y = 0$ are the minimum elements in A and B respectively, it follows that (0, 0) is the minimum element in the assessment space. Similarly, since $x = 100$ and $y = 100$ are the maximum elements in A and B respectively, it follows that (100, 100) is the maximum element in the assessment space (Figure 4.2).

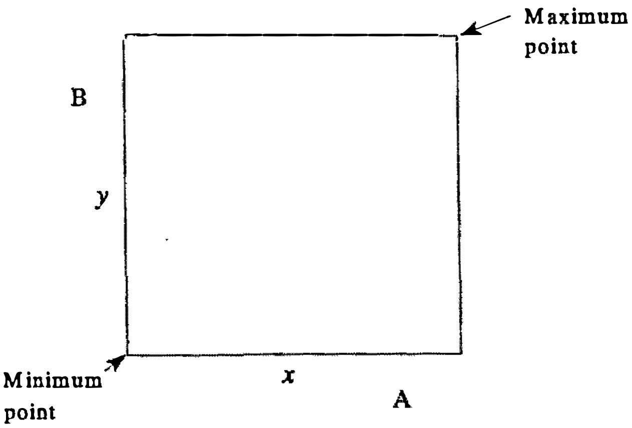


Figure 4.2: The Maximum point and Minimum point of the assessment space.

Equivalence

Equivalence, symbolised here as $\approx$, is defined to be the relationship in multidimensional space for which two elements, though different, i.e. $\neq$, are nevertheless not related by $\prec$. If two points from the assessment space, (x_1, y_1) and (x_2, y_2), are not equal, i.e. if $x_1 \neq x_2$, and $y_1 \neq y_2$, then exactly one of the following must be true:

1 $(x_1, y_1) \prec (x_2, y_2)$

2 $(x_2, y_2) \prec (x_1, y_1)$

3 $(x_1, y_1) \approx (x_2, y_2)$

Also, if $(x_1, y_1) = (x_2, y_2)$ then $(x_1, y_1) \approx (x_2, y_2)$, i.e., a point is equivalent to itself.

One way to represent equivalence in the assessment space is to use equivalence curves. An equivalence curve is a curve in the assessment space which represents the set of all points which are defined as being equivalent.

Equivalence curves will be symbolised here as γ. Figure 4.3 shows an equivalence curve in the assessment space. The two points r and s on γ are equivalent.

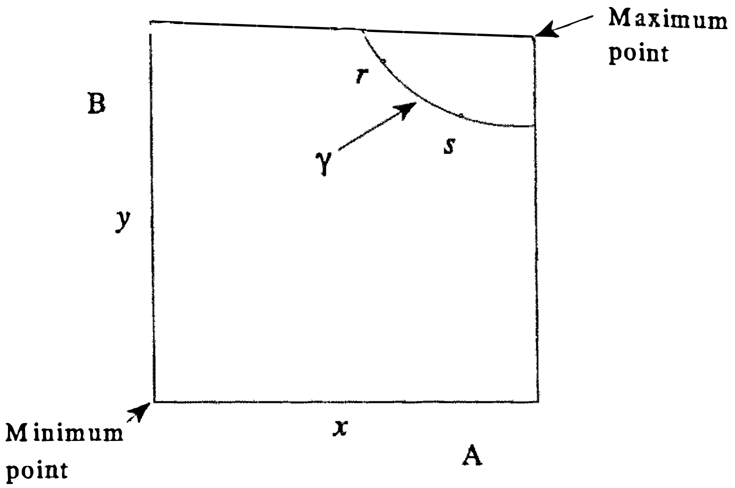


Figure 4.3: An equivalence curve γ in the assessment space.

Defining order in the assessment space

For this investigation the following definition will be used to define order in the assessment space:

For two points, a (x_a, y_a) and b (x_b, y_b), a is greater than b if $x_a > x_b$ and $y_a > y_b$

Target curve

Let there be a curve in the assessment space such that the minimum point and the maximum point lie on this curve. Call this curve the target curve and let it be symbolised as τ. Points on the target curve are ordered in the following way.

For two points on the target curve, c and d, if $c \neq d$ then $c \prec d$ if the distance of c to the minimum point along the target curve is less than the distance of d to the minimum point along the target curve. In Figure 4.4, c is less than d, i.e., $c \prec d$.

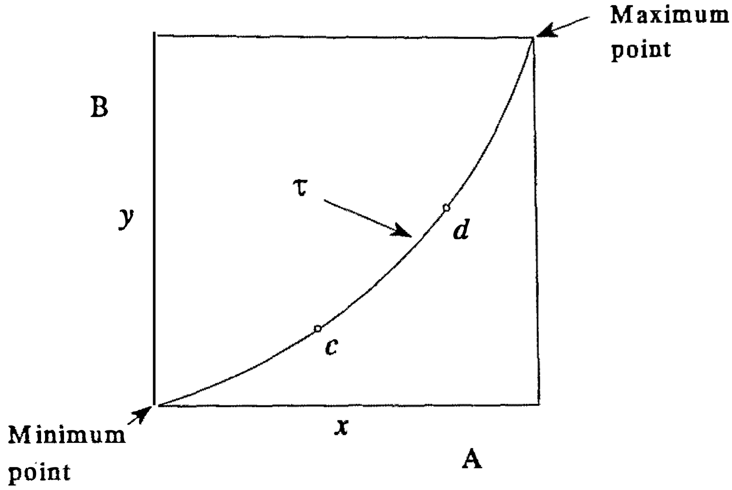


Figure 4.4: The target curve over the assessment space.

It can be shown that an equivalence curve intersects a target curve at exactly one place. Equivalence curves can therefore be ordered by ordering their points of intersection on the target curve.

Ordering mechanism (OMu)

The term **ordering mechanism (OMu)** refers to a unique family of equivalence curves and a target curve.

An important consequence of the definition of order in the assessment space is that if two equivalence curves are from the same family (ordering mechanism) then they are either the same curve or they do not intersect. That is, equivalence curves from the same family have the same shape. The proof of this is as follows.

If a family of equivalence curves consisted of curves with different shapes then it would be possible for at least two of these curves to intersect at some points and not at others (Figure 4.5). In Figure 4.5, the equivalence curves γ_1 and γ_2 intersect at point e. Since e lies on both γ_1 and γ_2, $f \approx e$ and $g \approx e$. Therefore $f \approx g$. Also, all of the points bounded by γ_1 and $_2\ \gamma$ are equivalent. In Figure 4.6, all of the points in the shaded region are equivalent.

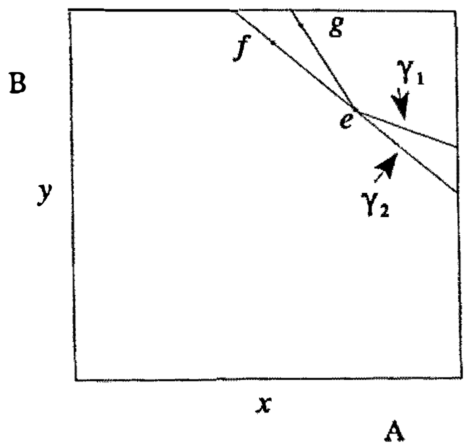


Figure 4.5: Intersecting equivalence curves.

If there are two dimensional regions of equivalence in the assessment space, then there will be points (x_1, y_1) and (x_2, y_2) such that although $x_1 > x_2$ and $y_1 > y_2$, $(x_1, y_1) \approx (x_2, y_2)$.

According to the definition of order in the assessment space used for this study this cannot occur. Hence, families of equivalence curves must be of the same shape.

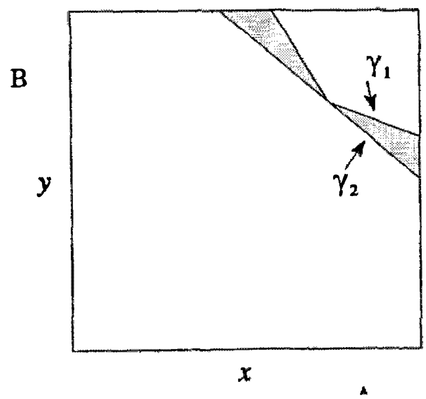


Figure 4.6: Equivalence regions

4.3 Developing seven ordering mechanisms

This section will develop seven different ordering mechanisms together with p-algorithms for each OMu . The following symbolism will be used in this section:

- OMui, where 1 ranges from 1 to 7, e.g., OMu1 refers to the first ordering mechanism. The numbers 1 to 7 are arbitrary and do not represent an order of priority.
- di refers to the distance algorithm derived from OMui. This algorithm expresses the mapping of a point onto the target curve as the distance of the intersection of the equivalence curve and the target curve from the minimum point on the target curve. d1, for example, is the distance algorithm for OMu1.
- pi refers to the percentage algorithm derived from di, and is therefore a linear transformation of di onto a scale from 0 to 100. The percentage algorithm expresses the value of di as a percentage of the total length of the target curve. p1, for example, is the percentage algorithm for OMu1. For this study, U is the scale onto which the percentage algorithms map the assessment data. It is a 100 point interval scale.

OMu1

OMu1 consists of a linear target curve and a family of linear equivalence curves such that

$\gamma: y = -x + c$

$\tau: y = x$

An illustration of OMu1 is shown in Figure 4.7.

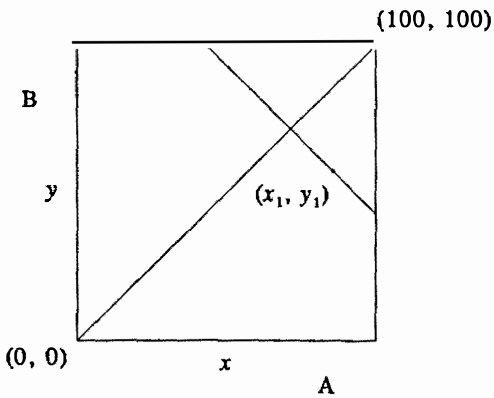


Figure 4.7: OMu1

For (x_1, y_1),

$\gamma: \quad y - y_1 = -(x - x_1) = x_1 - x$

$\tau: \quad y = x$

Solving γ and τ we obtain

$$y - y_1 = -(x - x_1) = x_1 - x$$
$$\therefore 2y = x_1 + y_1$$
$$\text{i.e., } y = \frac{x_1 + y_1}{2},\ x = \frac{x_1 + y_1}{2}$$

So,

$$\begin{aligned} d1 &= \sqrt{(x-0)^2 + (y-0)^2} \\ &= \frac{\sqrt{2(x_1 + y_1)^2}}{4} \\ &= \frac{(x_1 + y_1)}{\sqrt{2}} \end{aligned}$$

$$\text{and} \quad p1 = \frac{(x_1 + y_1)}{2}$$

OMu2

OMu2 consists of a linear target curve and a family of linear equivalence curves such that:

γ: $y = -mx + c$

τ: $y = x$

where m $\neq$ 1

A illustration of OMu2 is shown in Figure 4.8.

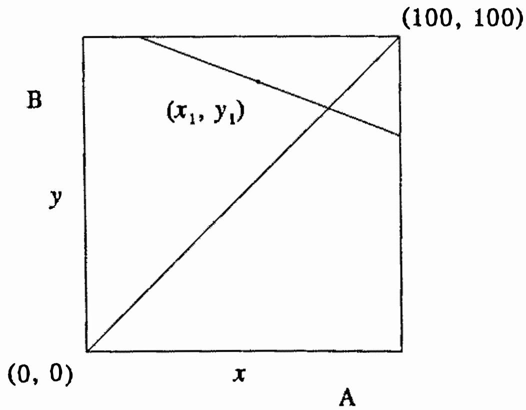


Figure 4.8: OMu2

For (x_1, y_1),

γ: $\quad y - y_1 = -m(x - x_1) = mx_1 - mx$

τ: $\quad y = x$

Solving γ and τ we obtain

$$y - y_1 = -m(y - x_1)$$
$$\therefore y - y_1 = -my + mx_1$$
$$\therefore y(1 + m) = y_1 + mx_1$$
$$\therefore y = \frac{mx_1 + y_1}{m + 1},\ x = \frac{mx_1 + y_1}{m + 1}$$

So,

$$d2 = \sqrt{2(\frac{mx_1 + y_1}{m + 1})^2}$$
$$= \sqrt{2}(\frac{mx_1 + y_1}{m + 1})$$

and $$p2 = \frac{mx_1 + y_1}{m + 1}$$

OMu3

OMu3 consists of a linear target curve and a family of circular equivalence curves centred on the minimum point (0, 0) such that:

γ: $x^2 + y^2 = a^2$

τ: $y = x$

A illustration of OMu3 is shown in Figure 4.9.

For (x_1, y_1),

γ: $x^2 + y^2 = x_1^2 + y_1^2$

τ: $y = x$

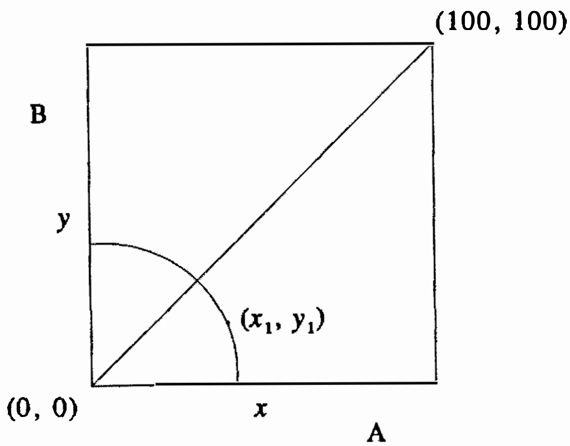


Figure 4.9: OMu3

Solving γ and τ we obtain

$$y^2 + y^2 = x_1^2 + y_1^2$$

$$\therefore 2y^2 = x_1^2 + y_1^2$$

$$\therefore y = \sqrt{\frac{x_1^2 + y_1^2}{2}}, \; x = \sqrt{\frac{x_1^2 + y_1^2}{2}}$$

So,

$$\text{d3} = \sqrt{2\frac{(x_1^2 + y_1^2)}{2}}$$

$$= \sqrt{x_1^2 + y_1^2}$$

$$\text{and p3} = \sqrt{\frac{x_1^2 + y_1^2}{2}}$$

OMu4

OMu4 consists of a linear target curve and a family of circular equivalence curves centred on the maximum point (100, 100) such that:

$$\gamma: (x - 100)^2 + (y - 100)^2 = a^2$$

$$\tau: y = x$$

A illustration of OMu4 is shown in Figure 4.10.

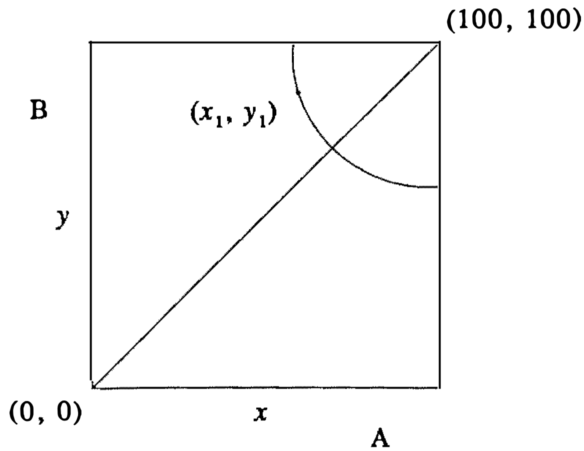


Figure 4.10: OMu4

For (x_1, y_1),

γ: $(x - 100)^2 + (y - 100)^2 = (x_1 - 100)^2 + (y_1 - 100)^2$

τ: $y = x$

Solving γ and τ we obtain

$$(y - 100)^2 + (y - 100)^2 = (x_1 - 100)^2 + (y_1 - 100)^2$$
$$\therefore 2(y - 100)^2 = (x_1 - 100)^2 + (y_1 - 100)^2$$
$$\therefore y - 100 = \pm\sqrt{\frac{(x_1 - 100)^2 + (y_1 - 100)^2}{2}}$$

Taking the negative value we obtain

$$y = 100 - \sqrt{\frac{(x_1 - 100)^2 + (y_1 - 100)^2}{2}}$$
$$= \frac{100\sqrt{2} - \sqrt{(x_1 - 100)^2 + (y_1 - 100)^2}}{\sqrt{2}}$$
$$\text{and } x = \frac{100\sqrt{2} - \sqrt{(x_1 - 100)^2 + (y_1 - 100)^2}}{\sqrt{2}}$$

So

$$d4 = \sqrt{\frac{2(100\sqrt{2} - \sqrt{(x_1 - 100)^2 + (y_1 - 100)^2})^2}{2}}$$
$$= 100\sqrt{2} - \sqrt{(x_1 - 100)^2 + (y_1 - 100)^2}$$
$$\text{and } p4 = \frac{100\sqrt{2} - \sqrt{(x_1 - 100)^2 + (y_1 - 100)^2}}{\sqrt{2}}$$

Omu5

OMu5 consists of a linear target curve and a family of rectangular equivalence curves such that:

$\gamma: y = c$ when $x_1 > y_1$ and $x = c$ when $y_1 > x_1$, where c is a constant

$\tau: y = x$

A illustration of OMu5 is shown in Figure 4.11.

For (x_1, y_1) where $y_1 > x_1$

γ: $x = x_1$

τ: $y = x$

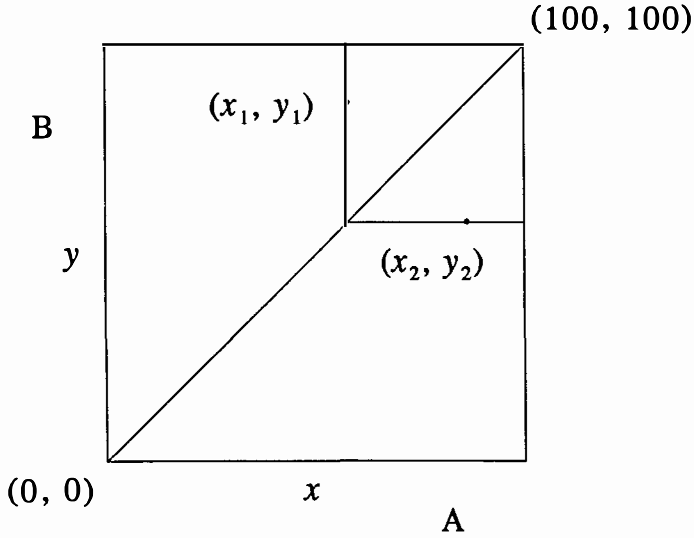


Figure 4.11: OMu5

Solving γ and τ we obtain $y = x = x_1$

So,

$$d5 = \sqrt{x_1^2 + x_1^2}$$
$$= x_1\sqrt{2}$$

and $$p5 = x_1$$

For (x_2, y_2) where $x_2 > y_2$

γ: $y = y_2$

τ: $y = x$

Solving γ and τ we obtain $y = x = y_2$

So,

$$d5 = \sqrt{y_2^2 + y_2^2}$$
$$= y_2\sqrt{2}$$
and $$p5 = y_2$$

OMu6

OMu6 consists of a linear target curve and a family of angular equivalence curves such that:

γ: $y = -kx + c$ when $x_1 > y_1$ and $y = -hx + c$ when $y_1 > x_1$

τ: $y = x$

A illustration of OMu6 is shown in Figure 4.12.

For (x_1, y_1) where $y_1 > x_1$

γ: $y = -kx + c$

τ: $y = x$

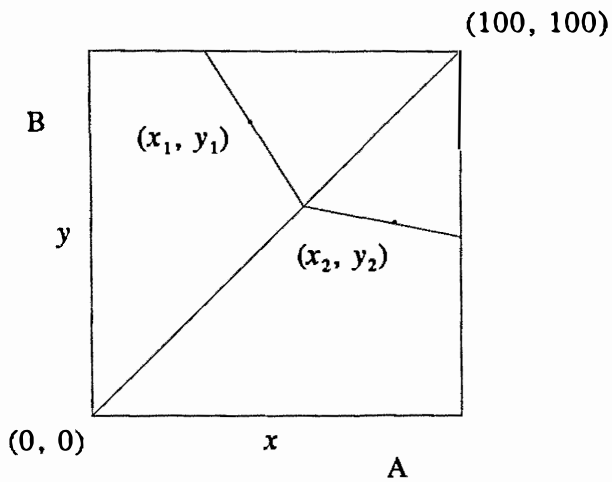


Figure 4.12: OMu6

Using the result from OMu2,

$$d6 = \sqrt{2}\left(\frac{kx_1 + y_1}{k + 1}\right)$$

and $$p6 = \frac{kx_1 + y_1}{k + 1}$$

For (x_2, y_2) where $x_2 > y_2$

γ: $y = -hx + c$

τ: $y = x$

Using the result from OMu2,

$$d6 = \sqrt{2}\left(\frac{hx_2 + y_2}{h + 1}\right)$$

$$\text{and} \quad p6 = \frac{hx_2 + y_2}{h + 1}$$

OMu7

OMu7 consists of a linear target curve and a family of equivalence curves such that:

γ: $(x - f)^2 + (y - g)^2 = (x_1 - f)^2 + (y_1 - g)^2$

τ: $y = x$

A illustration of OMu7 is shown in Figure 4.13.

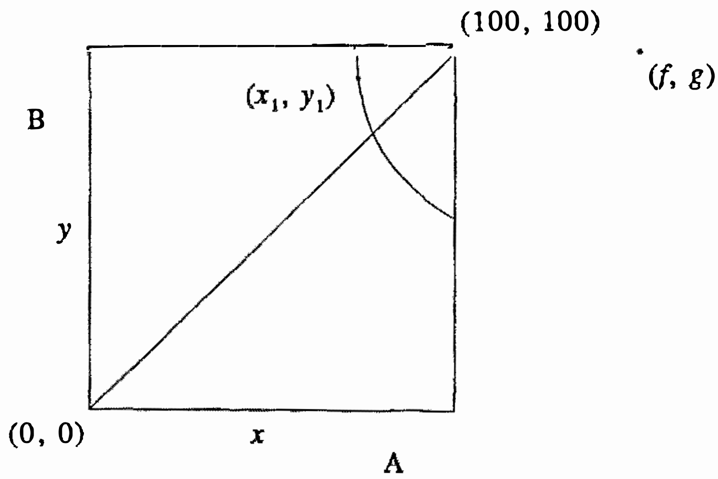


Figure 4.13: OMu7

Solving γ and τ we obtain

$$(y - f)^2 + (y - g)^2 = (x_1 - f)^2 + (y_1 - g)^2$$

$$\therefore \quad y^2 - 2fy + f^2 + y^2 - 2gy + g^2 = (x_1 - f)^2 + (y_1 - g)^2$$

$$\therefore \quad 2y^2 - 2(f + g)y + f^2 + g^2 = (x_1 - f)^2 + (y_1 - g)^2$$

$$\therefore \quad 2y^2 - 2(f + g)y + f^2 + g^2 - (x_1 - f)^2 - (y_1 - g)^2 = 0$$

This is a quadratic equation in y.

$$\therefore \quad y = \frac{2(f+g) - \sqrt{4(f+g)^2 - 8f^2 - 8g^2 + 8(x_1 - f)^2 + 8(y_1 - g)^2}}{4}$$

$$= \frac{2(f+g) - 2\sqrt{(f+g)^2 - 2f^2 - 2g^2 + 2(x_1 - f)^2 + 2(y_1 - g)^2}}{4}$$

$$= \frac{f + g - \sqrt{(f+g)^2 - 2f^2 - 2g^2 + 2(x_1 - f)^2 + 2(y_1 - g)^2}}{2}$$

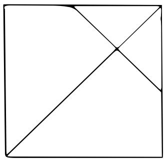
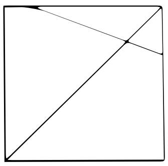
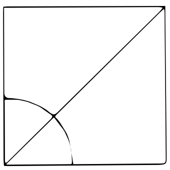
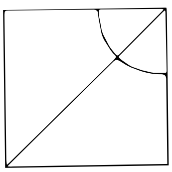
$$\text{So} \quad d7 = \frac{f + g - \sqrt{(f+g)^2 - 2f^2 - 2g^2 + 2(x_1 - f)^2 + 2(y_1 - g)^2}}{\sqrt{2}}$$

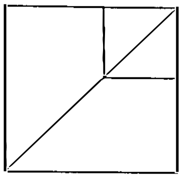
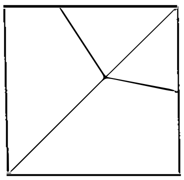
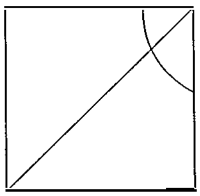
$$p7 = \frac{f + g - \sqrt{(f+g)^2 - 2f^2 - 2g^2 + 2(x_1 - f)^2 + 2(y_1 - g)^2}}{2}$$

Summary

The following table (Table 4.1) summarises the percentage algorithms for each of the ordering mechanisms for ordering points (x, y) in the assessment space.

Table 4.1
Summary of percentage algorithms for OMu

OMu1		$$p1 = \frac{(x + y)}{2}$$
OMu2		$$p2 = \frac{(mx + y)}{m + 1}$$
OMu3		$$p3 = \sqrt{\frac{x^2 + y^2}{2}}$$
OMu4		$$p4 = \frac{100\sqrt{2} - \sqrt{(x - 100)^2 + (y - 100)^2}}{\sqrt{2}}$$

OMu5		For (x, y) where $x > y$ $$p5 = y$$ For (x, y) where $y > x$ $$p5 = x$$
OMu6		For (x, y) where $x > y$ $$p6 = \frac{kx + y}{k + 1}$$ For (x, y) where $y > x$ $$p6 = \frac{hx + y}{h + 1}$$
OMu7		$$p7 = \frac{f + g - \sqrt{(f + g)^2 - 2f^2 - 2g^2 + 2(x_1 - f)^2 + 2(y_1 - g)^2}}{2}$$

The OM which are summarised in Table 4.1 can be interpreted as representing decision structures for the overall value of a profile of achievement for a percentage-based assessment system consisting of 2 criteria. The interpretation of each of the OM in Table 4.1 will now be addressed.

OMu1

OMu1 is an ordering mechanism which is indifferent to the arrangement of achievement within a set of criteria. For example, the result (50, 100) is equivalent to (100, 50), or more generally, $(a, b) \approx (b, a)$. This OM is symmetrical about the target curve and weights each of the criteria equally, considering achievement in one criterion to be as valuable to overall achievement as achievement is the other criterion.

The decision structure for this OM is one for which overall achievement is determined quantitatively rather that qualitatively, in that the arrangement of achievement across the criteria is irrelevant to the overall achievement. Also, there is 'equal tradeoff' on both criteria, so that a strength in one criterion is fully compensated by an equivalent weakness in the other criterion. Also, OMu1 treats perfect achievement in one criterion and no achievement in the other, (100, 0) or (0, 100), to be equivalent to 'half-marks' in both, i.e., (50, 50).

OMu2

OMu2 is an ordering mechanism which allows for a weakness in one criterion to be compensated by a strength in another, similar therefore to OMu1, but for OMu2 achievement in one criterion is more important for overall achievement than achievement in the other (since $m \neq 1$). For results where $a \neq b$, $(a, b) \not\approx (b, a)$. Clearly, OMu2 is a weighted average and is not symmetrical about the target curve.

The decision structure for OMu2 is one which allows for tradeoff, but tradeoff favouring one criterion over another.

Omu3

OMu3 is an ordering mechanism which defines overall achievement to be the Euclidean distance from the minimum. Achievement is represented as a position vector and overall achievement is defined as the magnitude of this position vector. What this means in terms of a decision structure, however, is that a strength in one criterion at the expense of an equivalent weakness in the other criterion is rewarded. For example, the result (100, 0), or (0, 100) is considered to be better than the result (50, 50), and equivalent to the result (70, 70). Also, tradeoff for OMu3 is symmetrical about the target curve, i.e., $(a, b) \approx (b, a)$.

OMu4

OMu4 defines overall achievement as the Euclidean distance from the maximum. The decision structure is one which allows for symmetrical tradeoff about the target curve, but one which penalises a weakness in one criterion against a strength in the other. This is the opposite to OMu3. Also, an increase in overall achievement by a fixed increase in achievement in one criterion is sensitive to the location of the result in the assessment space. For example, if the result lies above the target curve, then maximum benefit in terms of overall achievement will be obtained by an increase of a fixed amount in the criterion along the horizontal axis.

OMu5

OMu5 is the no-tradeoff ordering mechanism. This OM formalises for the entire assessment space decision structures represented by decision tables specified in terms of minimum cutoffs in each of the criteria. For example, suppose a VHA and a HA are defined according to Table 4.2.

Table 4.2
Cutoffs for VHA and HA

	x	y
VHA	≥ 80	≥ 80
HA	≥ 60	≥ 60

A result of (81, 90) is clearly a VHA. A result of (79, 100) is a HA. The boundary of the VHA set defines the equivalence curve for the assessment space. As acknowledged by Sadler (1988, p.10; see section 1.5) the OM to represent this decision structure cannot be any form of additive composition rule. As shown above, OMu5 defines a student's overall achievement in terms of the lowest result in any of the criteria, irrespective of the other results in the profile.

Some problems with the no-tradeoff decision structure can be illustrated using OMu5. Consider two students, *r* and *s*, whose results are shown in the assessment space in Figure 4.14.

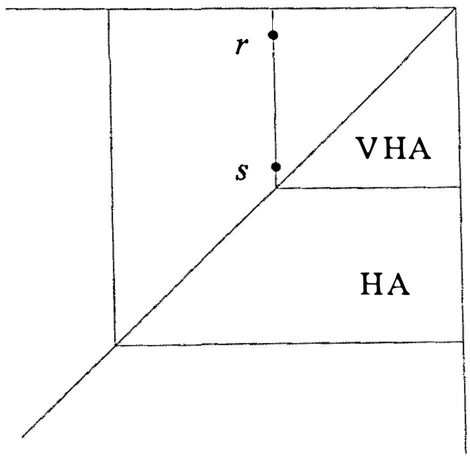


Figure 4.14: Equivalent borderline VHA

The no-tradeoff decision structure results in *r* and *s* being equivalent, despite *r* being equal to *s* in one criterion and greatly exceeding *s* in the other. If *r* were to shift slightly down the *x*-criterion, *r* would be less than *s*, since *r* would be a HA (Figure 4.15), and would therefore have to be ranked lower than *s* on the univariate scale since a lower LOA must be awarded a lower SAI, (BSSSS, 1992) despite being much greater than *s* on the other criterion.

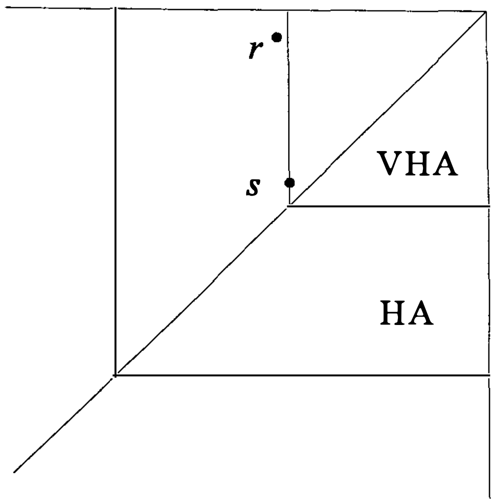


Figure 4.15: Student *r* lower than *s*

If no-tradeoff decision tables are used to first group results into LOA categories and then once in the category, a different decision structure is used to rank results within an LOA (e.g. OMuVHA), some anomalies arise. Such an OS (a 'category' OS) is represented in Figure 4.16.

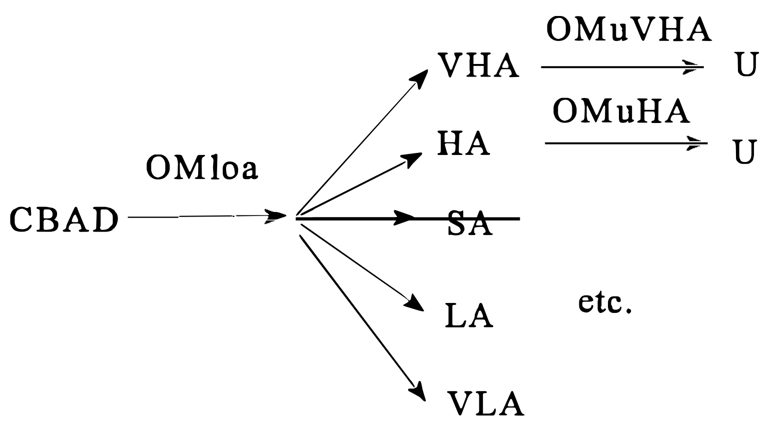


Figure 4.16: A category OS

Such an OS results in different shaped equivalence curves in the assessment space. Consider the situation where OMloa is OMu5 such that VHA and HA are defined as in Table 4.2 and where OMuVHA is OMu1 (i.e., an unweighted average of the results in the VHA band). Once in the VHA category, a point is mapped onto the target curve using a linear equivalence curve with gradient of -1. The equivalence curves represented by the boundary between HA and VHA, together with the equivalence curve for ordering points in the VHA category, are shown in Figure 4.17.

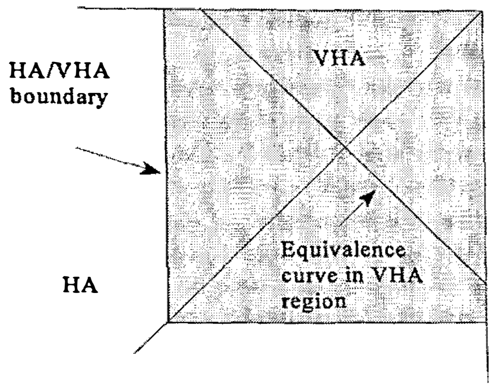


Figure 4.17: Two different equivalence curves in a category OS

Suppose *r* is (79, 100), *s* is (80, 100) and *t* is (80, 80). According to the decision structure underlying this ordering system, $r < t < s$ (Figure 4.18). In the assessment space, *r* and *s* differ by 1% in one criterion yet the difference between them is approximately 10% on the univariate scale. The difference in overall achievement does not seem to reflect the differences among the three scores and particularly the differences between *r* and *s*. It is worth noting that, under this combination of curves, there is no way to end up between 80 and 90 on the scale U.

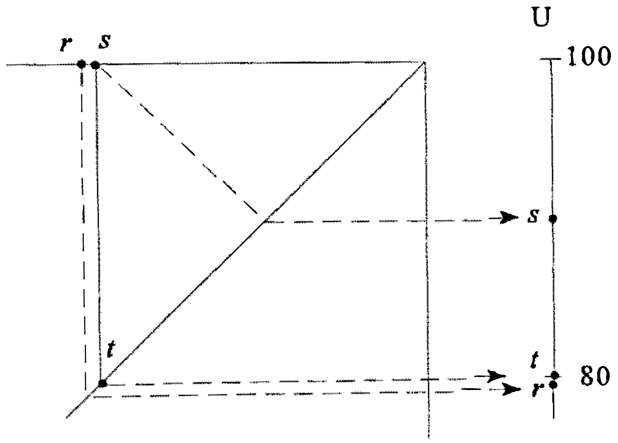


Figure 4.18: Differences in data not reflected using category OS.

Omu6

OMu6 is an ordering mechanism which represents a relaxation of the stringent no-tradeoff decision structure behind OMu5. OMu6 showS how it is possible to have different 'rates' of tradeoff on different criteria, i.e.,

(ii) allowable tradeoff on different criteria, and

(iii) non-negotiable minima on some or all criteria (Sadler, 1988, p.10).

OMu7

OMu7 is an ordering mechanism which allows for tradeoff between criteria but tradeoff is non-symmetrical about the target curve. This ordering mechanism combines ideas from OMu2 and OMu4 such that there is a difference in weighting of different criteria, reflecting a valuing of performance in one criterion over performance in the other.

The significance of the target curve.

The seven ordering mechanisms developed above share a common target curve and differ only in the shape of the equivalence curves. The target curve is of the form $y = x$. There are, however, many different target curves available for generating OMu. OMul to OMu7 can be classified as a family of ordering mechanisms under a common target curve ($y = x$) and may therefore be called a 'target curve family of ordering mechanisms'. Other ways of classifying ordering mechanisms is to have a common equivalence curve and different target curves. On this basis it would be possible to generate an entire catalogue of ordering mechanisms grouped according to a common target curve and with different equivalence curves, or a common equivalence curve and different target curves. This could form the basis of a systematic development and study of the properties of a large number of ordering mechanisms, a task, however, beyond this study.

(The determination of d-algorithms and p-algorithms is slightly more complicated for target curves not of the form $y = mx$, since the distance of a point along the target curve from the minimum involves determining the length of a segment of curve. Consider the point r (x_1, y_1) on the target curve τ: $y = f(x)$ in Figure 4.19.

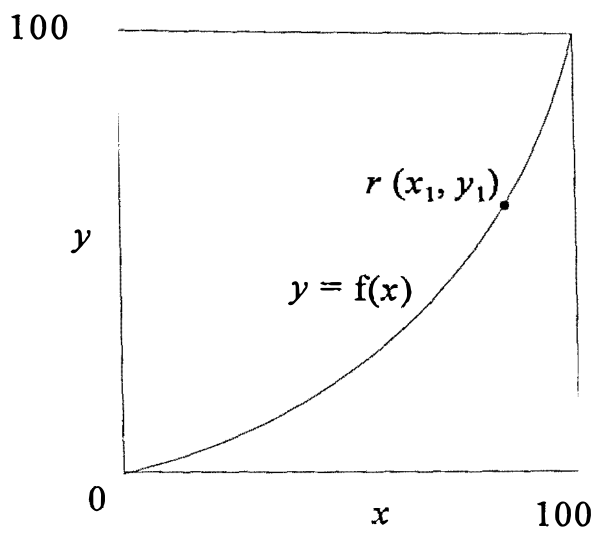


Figure 4.19: Point r lying on a non-linear target curve

For point r, $d = \int_0^{x_1} \sqrt{1 + f'(x)}\, dx$ and $p = \dfrac{\int_0^{x_1} \sqrt{1 + f'(x)}\, dx}{\int_0^{100} \sqrt{1 + f'(x)}\, dx} \times 100$).

One of the fundamental differences between the ordered distributions generated by each of the OM developed in section 4.3 is the relative overall position on U of points off the target curve with respect to points on the target curve. All points on the target curve are awarded the same p-value by each of the OMu1 to OMu7.

Consider two points, r and s shown below (Figure 4.20).

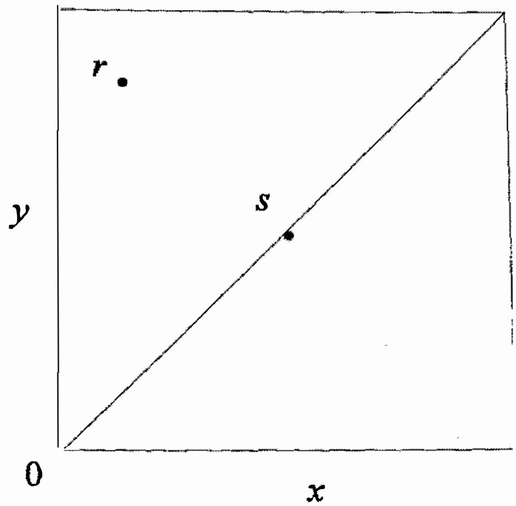


Figure 4.20: Two points equivalent according to OMu1

Suppose r and s are equivalent according to OMu1, i.e., OMu1: $r \approx s$. Now, OMu2: $s > r$, OMu3: $r > s$, OMu4: $s > r$, OMu5: $s > r$, OMu6: $s > r$ (depending, of course, on the values of k and h), OMu7: $s > r$. Five of the seven OMu rank s higher than r, one of the seven rank s lower than r, and one ranks s and r equally. There is no absolute unambiguous overall order for multivariate data since the data do not form an ordered set (section 2.4.2). The question 'is $r > s$, $r \approx s$, or $r < s$?' is always defined in the end by a decision structure which imposes an order on the data. Ordering mechanisms, of the kind developed here, do not replace decision structures but simply attempt to model them. Calculations of overall achievement are really models of decision structures. Hence, SAIs should be able to be both a calculation and a decision (BSSSS, 1992, p. 3).

If a cohort of students achieve results which lie along the target curve then the differences between the ordered distributions for each of the ordering mechanisms will be minimal. In other words, if achievement in one criterion influences achievement in the other criterion, then results should increase similarly across the criteria, meaning the likelihood of extreme results is reduced and that the cohort should fall around the target curve. The choice of ordering mechanism, although having some effect on the ordered distribution, does not have as much bearing on the distribution compared with the situation for which results are scattered around the target curve.

The differences between the ordered distributions for OMu1 to OMu7 will now be investigated using computer simulation.

4.4 Computer simulated investigations

This section will discuss the results from six computer simulation investigations, labelled investigation 1 through to investigation 6. Investigation 1, 2 and 3 involve the seven 2-d p-algorithms developed in 4.3. The constants for OMu6 used for these investigations are $k = 2$ and $h = 0.5$. The constants for OMu7 are $f = 200$, $g = 100$, and for OMu2, $m = 2$. Investigation 4, 5 and 6 are for 3-d p-algorithms derived from the 2-d algorithms for OMu1, OMu3, OMu4 and OMu5. These three investigations demonstrate the generalisation and subsequent use of p-algorithms to more than 2 dimensions. This is a preliminary investigation to develop and apply the 3-d algorithms which are then used in investigation 7 (section 4.5) to design a set of independent OS to order percentage-based CBAD for a cohort of senior chemistry students.

4.4.1 Investigation 1

Investigation 1 is a computer simulated investigation of the ordered distributions of 2-d percentage based data for which 100 data points are generated randomly throughout the entire 2-d assessment space. The program for the simulation is in appendix B and the table of data generated from one run of the program is in appendix C.

A scatterplot of the assessment data generated for investigation 1 is shown in Figure 4.21.

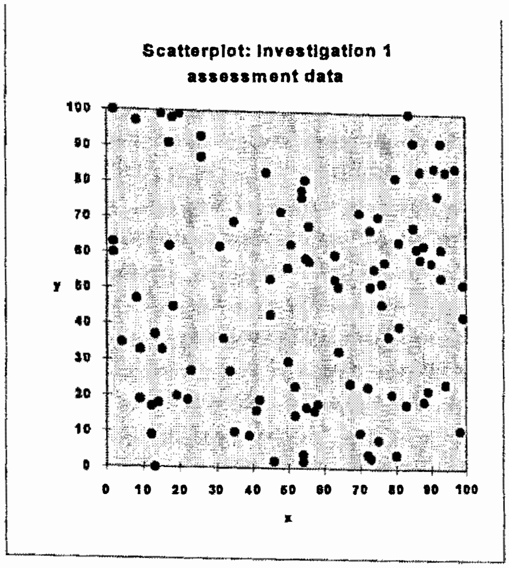


Figure 4.21

This simulation represents a situation for which achievement in one criterion is independent of achievement in the other; hence, the random distribution of points throughout the assessment space since there is no correlation.

The correlation coefficient for each of the distinct pairs of distributions is shown in Table 4.3.

Table 4.3

Correlation coefficients for pairs of p-values from investigation 1

	p1	p2	p3	p4	p5	p6	p7
p1	1						
p2	0.953	1					
p3	0.974	0.925	1				
p4	0.973	0.934	0.900	1			
p5	0.871	0.831	0.742	0.954	1		
p6	0.981	0.935	0.915	0.997	0.948	1	
p7	0.902	0.986	0.854	0.907	0.828	0.902	1

The least correlated pair of distributions are p3 and p5 (r = 0.742). A scatterplot of p3 versus p5 is shown in Figure 4.22.

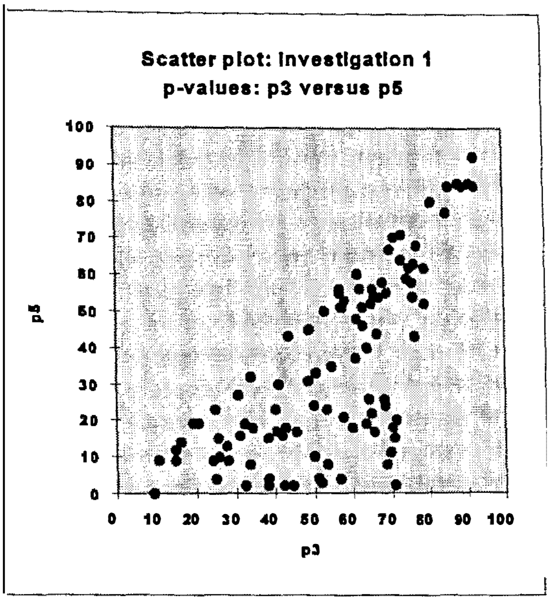


Figure 4.22

The scatterplot allows for a visual interpretation of the similarities and differences between these two OMs. A straight line drawn across the scatterplot from (0, 0) to (100, 100) will be called the 'agreement line' since it represents those points in the assessment space which are awarded the same p-value by each OM. The points which lie along the target curve in the assessment space are mapped onto the agreement line in each scatterplot. Those points which lie below the agreement line are awarded a higher p-value by the OM along the horizontal axis; those points which lie above the agreement line are awarded a higher p-value by the OM along the vertical axis. Clearly, for the scatterplot in Figure 4.22, OMu3 is awarding higher p-values to points off the target curve than OMu5. This is not surprising since OMu5 represents a no-tradeoff decision structure where each point is ordered by its lowest component, and OMu3 favours points off the target curve with respect to points on the target curve.

Extreme differences between OMu3 and OMu5 are evident for certain points. Point 84 (see appendix C) has the result (2, 100). This point lies just to the right of 70 along the p3-axis in Figure 4.22. For point 84, p3 = 71 whereas p5 = 2. This is an overall difference of 69% (note: since U is a scale from 0 to 100 the difference of 1 p-value is a 1% difference on the U scale). The choice of ordering mechanism makes a significant difference to the overall position on U of point 84. For other points, e.g., point 9 (56, 58), where p3 = 57 and p5 = 56, the different OMs produce a negligible difference in p-values. OMu3 and OMu5 generate different ordered distributions and the differences between these two OM are manifest for points which lie away from the target curve.

The greatest correlation coefficient is for p4 versus p6 (r = 0.997). The scatterplot for p4 versus p6 is shown in Figure 4.23. Notice how the points in the scatterplot cluster around the agreement line. This can be explained by the similar geometry of the equivalence curves for these two ordering mechanisms (compare the equivalence curves for these ordering mechanisms in Table 4.1).

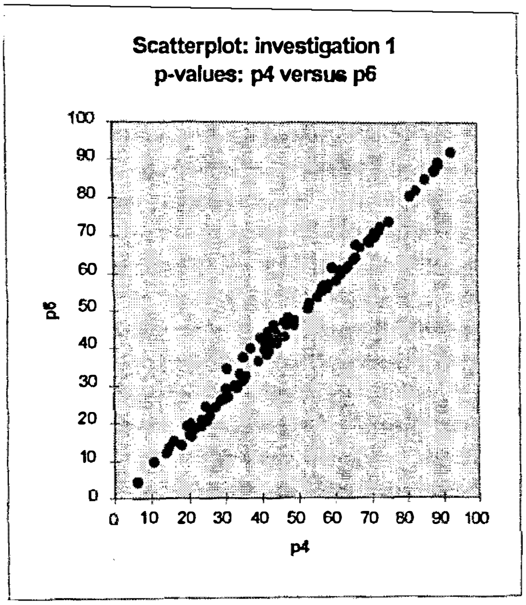


Figure 4.23

A complete discussion of all pairs of p-values is beyond this study. What will be addressed for the remainder of this investigation, however, is how p2 to p7 compare with p1. This will give some idea of the similarities and differences between the p-values awarded by each of the OM, and will enable some comparasion of the preferences represented by each OM (research question part 2).

A bar chart showing the correlation coefficient between OMu1 and the other six ordering mechanisms is shown in Figure 4.24.

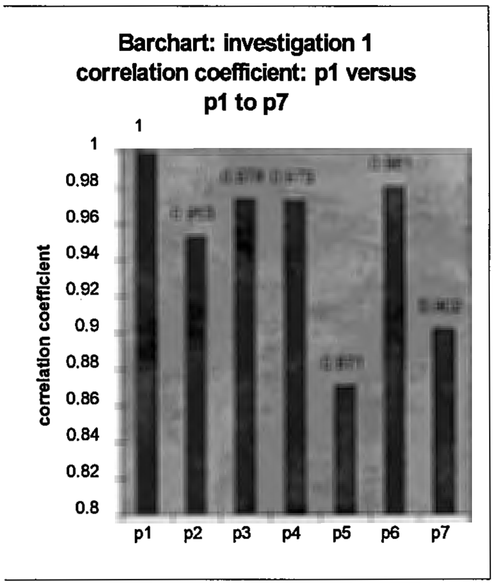


Figure 4.24

The least degree of correlation is between p1 versus p5 (r = 0.871). The scatterplot for p1 versus p5 is shown in Figure 4.25.

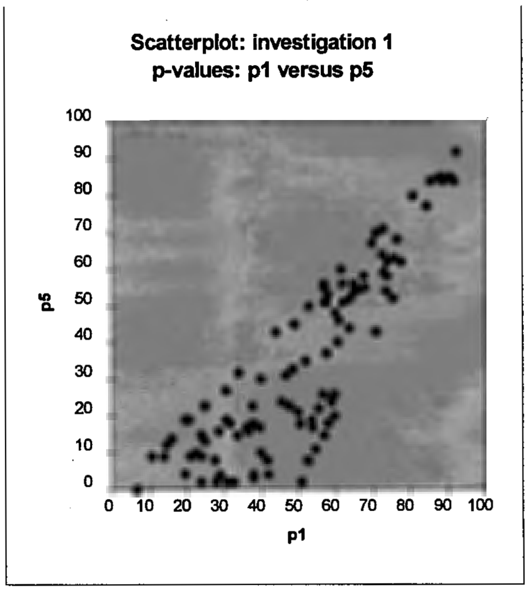


Figure 4.25

This scatterplot is similar to that for p3 versus p5 (Figure 4.22). The points in this scatterplot lie either on or below the agreement line; those points which lie near the agreement line in this scatterplot are results which lie near the target curve in the assessment space.

The differences between OMu1 and OMu5 (Figure 4.25) are those differences which would be expected between and averaging decision structure (OMu1) and a no-tradeoff, or minimum achievement decision structure (OMu5). An averaging decision structure would map students with widely varying results across the 2 criteria higher onto the univariate scale than a no-tradeoff decision structure. The bulk of the points in the scatterplot in Figure 4.25 lie below the agreement line, indicating that p1 is indeed mapping points off the target curve higher on U than p5. Hence, the p-algorithms appear to exhibit properties consistent with the decision structures which they are believed to model. OMu5 clearly prefers points where the components are at similar levels over points where components are at different levels.

The order of the ordering mechanisms in terms of correlation with OMu1 from lowest to highest is OMu5 (r = 0.871), OMu7 (5 = 0.902), OMu2 (r = 0.953), OMu4 (r = 0.973), OMu3 (r = 0.974) and OMu6 (0.981). The p-value scatterplots for each of these with p1 is shown below (Figure 4.26 to Figure 4.30).

It is clear from these scatterplots that these ordering mechanisms are generating inherently different ordered distributions . The OM are not agreeing on the order nor the gaps between the 2-d percentage-based data. Only those points on the target curve are awarded the same p-value by each of the 7 ordering mechanisms. All other points are awarded different p-values. These 7 OM are 'saying' different things about the overall value, and in particular, the relative overall value, of the profile of results for each data point in the set. Different ordered distributions are obtained for the same set of data. The OMu are representing different preferences in the assessment space.

Some patterns can be identified from the scatterplots in Figure 4.26 to 4.30. OMu4 and OMu6 generate the most similar looking scatterplots with OMu1 (Figure 4.28 and Figure 4.30 respectively). In both cases, OMu1 is awarding higher p-values to points above or below the target curve than either OMu4 or OMu6, since the points in the scatterplots lie on or below the agreement line. There are no points above the agreement line. OMu3, like OMu4 and OMu6, have points on or to the one side of the agreement line, but for p1 versus p3, the points lie above the agreement line indicating that OMu3 is awarding higher p-values to points above or below the target curve in the assessment space.

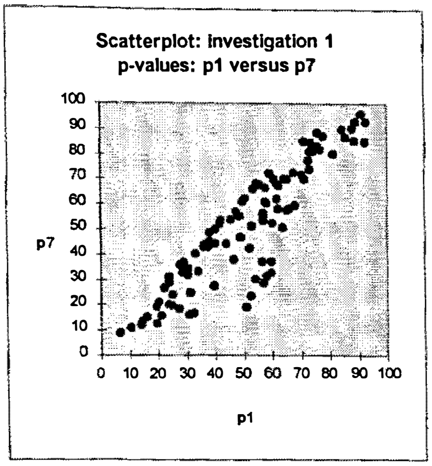


Figure 4.26

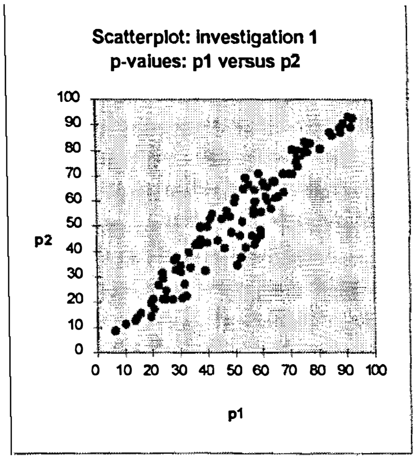


Figure 4.27

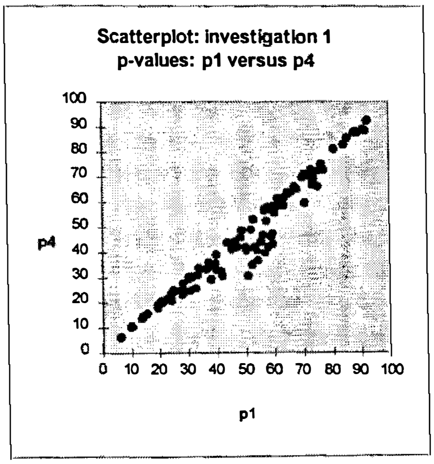


Figure 4.28

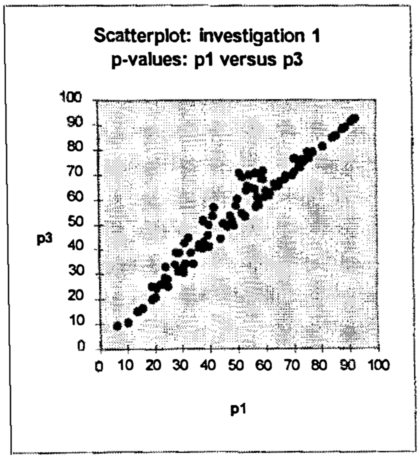


Figure 4.29

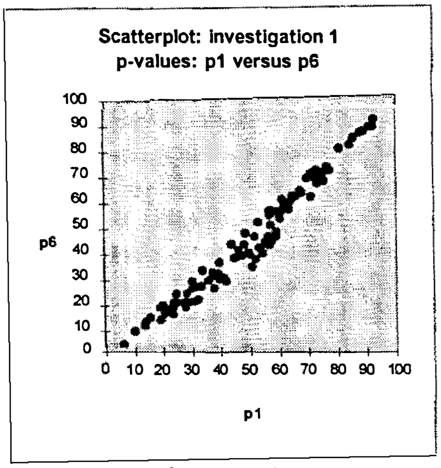


Figure 4.30

Like OMu4 and OMu6, OMu5 is awarding lower p-values to points off the target curve, but the differences between OMu5 and OMu1 (Figure 4.22) is much greater than the differences between OMu4 and OMu1, and OMu6 and OMu1. In the context of educational assessment, the differences between OMu5 and OMu1 would be controversial. A student with the result (2, 100) (point 84, appendix B) is awarded a p-value of 51 by OMu1 and 2 by OMu5. This student would perhaps drop from an SAI of around 300 to an SAI near 200, and from an SA to a VLA. Now OMu1 is an aggregating decision structure and OMu5 a no-tradeoff decision structure. It is interesting to note that the no-tradeoff decision structure is argued by some (Hewitson, 1988) to be the only decision structure compatible with criterion-referenced (criteria-based) assessment.

The equivalence curves for OMu1, OMu3, OMu4, OMu5 and OMu6 (for the values of *k* and *h* chosen for these investigations) are symmetrical about the target curve. This is not the case for OMu2 and OMu7. The non-symmetrical nature of these two OM is manifest in the scatterplots with respect to OMu1 (Figure 4.27 and 4.26 respectively). These scatterplots show that there are points lying either side of the agreement line, something not seen in the p-value scatterplots for the symmetrical ordering mechanisms.

Consider the scatterplot for p1 versus p2 (Figure 4.27). It can be shown that OMu2 is favouring points which lie below the target curve, i.e., points (x, y) for which $x > y$, and these points are mapped above the agreement line in the scatterplot. For example, point 12 (72, 4) (appendix C) lies below the target curve and for this point p1 = 38 and p2 = 49. Points above the target curve are favoured by OMu1 with respect to OMu2, i.e., points (x, y) where $y > x$, and are mapped below the agreement line. Now, OMu1 equivalence curves are symmetrical about the target curve, whereas the equivalence curves for OMu2 are non-symmetrical. Figure 4.31 shows the target curve and two equivalence curves, one for OMu1 (γ_1) one for OMu2 (γ_2).

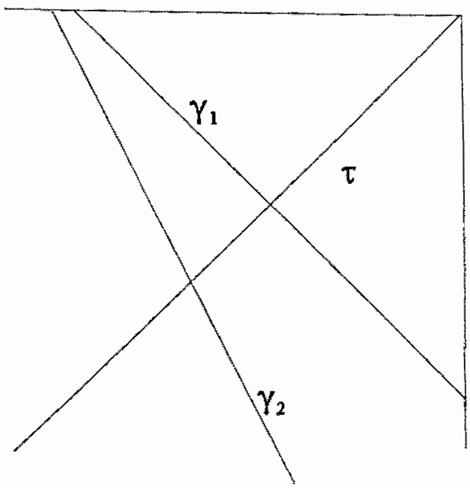


Figure 4.31: The target curve, γ_1 and γ_2

Consider point *r* which lies below the target curve. This point is mapped to position 1 on the target curve by OMu1, and to position 2 by OMu2 (Figure 4.32). 1 lies lower on the target curve than 2 (i.e. 1 is further from the max than 2), hence p1 for *r* is lower than p2.

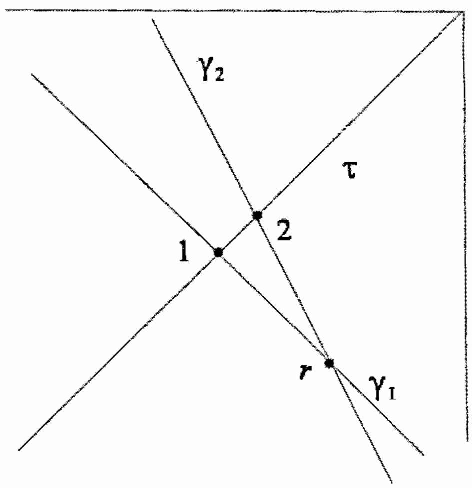


Figure 4.32: Point *r* mapped onto the target curve by γ_1 and γ_2.

For point *s* which lies above the target curve, however, the OMu1 equivalence curve maps *s* higher than OMu2 (Figure 4.33). Hence p2 for *s* is lower than p1.

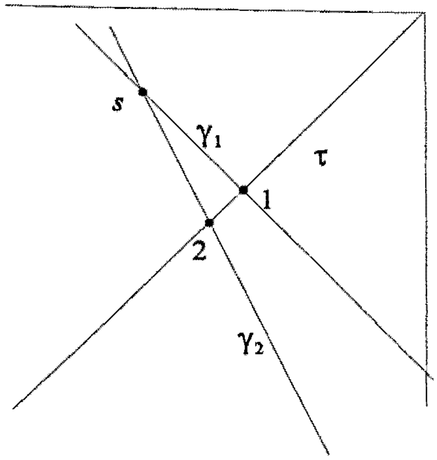


Figure 4.33: Point s mapped onto the target curve by γ_1 and γ_2.

OMu7 compares with OMu1 in a similar way OMu2 compares with OMu1. OMu7 weights the horizontal axis criterion more heavily than the vertical. The points below the agreement line are those points above the target curve in the assessment space; the points above the agreement line are those points below the target curve. Whilst OMu7 ranks points below the target curve in a similar way to OMu2 with respect to OMu1, OMu7 ranks extreme points above the target curve lower than OMu2. For example, point 82 (2, 100) (appendix C) is awarded a p-value of 35 by OMu2 and 19 by OMu7.

At this stage it is possible to make two statements from investigation 1. These are labelled 1 and 2 for future reference.

1 The seven OMu developed in section 4.3 from the mathematical system developed in section 4.2, generate a p-value, and hence a position on an ordered scale, for every point generated in the assessment space. It can be shown that the seven OMu order all points in the assessment space.

This finding directly responds to the first part of the research question, i.e., is there a mathematical system which orders all points in a multidimensional assessment space? The mathematical system developed in section 4.2 is such a system.

2 The seven OMu generate inherently different rank order distributions for data generated randomly throughout the assessment space. These differences occur in the way each of the OMu assign a p-value to points off the target curve with respect to points lying on the target curve. That is:

(i) Points off the target curve are awarded different p-values by each OMu. The further the points from the target curve, the greater the difference in p-values awarded by each of the OMu.

(ii) Points on the target curve are awarded the same p-values by each of the seven OMu, and points in close proximity to the target curve are awarded similar p-values by each of the OMu.

The differences between the p-values awarded by each of the OM represent different preferences for regions in the assessment space. Hence, this finding directly responds to the second part of the research question, i.e., is there a mathematical system which models preferences within a miltidimensional percentage-based assessment space? The mathematical system developed in section 4.2 seems to be such a system.

The following two investigations, investigation 2 and investigation 3 will explore further 2(i) and 2(ii) respectively.

4.4.2 Investigation 2

Investigation 2 is a computer simulated investigation of the ordered distribution of 2-d percentage-based data for which 100 data points are generated randomly within two confined regions of the assessment space. The first region is a 'non-extreme' region around the centre of the assessment space where $40 \leq x \leq 60$ and $40 \leq y \leq 60$. The non-extreme regions of the assessment space are those regions that lie around the target curve. The second region is an 'extreme' region where $0 \leq x \leq 20$ and $80 \leq y \leq 100$. Extreme regions lie away from the target curve, i.e., at the extremeties of the assessment space. The program used for this investigation is in appendix D and the data generated from one run of the program is in appendix E.

This investigation is exploring point 2(i) made in section 4.4.1, investigation 1, that:

> Points off the target curve are awarded different p-values by each OMu. The further the points from the target curve, the greater the difference in p-values awarded by each of the OMu.

Such differences can be described as preferences for regions in the assessment space and are of relevance to the second part of the research question.

The assessment data generated for this investigation are shown in the scatterplot in Figure 4.34. The data in the extreme region are in the top left corner of the assessment space, and the data in the non-extreme region are in the centre of the assessment space.

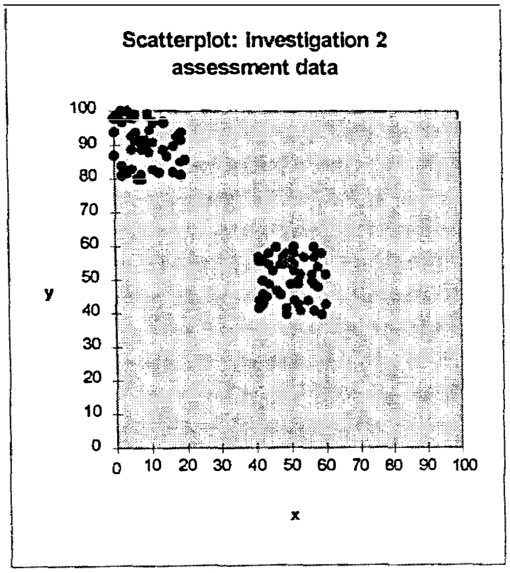


Figure 4.34

The correlation coefficients have been calculated for the p-values for each pair of OMu (Table 4.4). At a glance there is a wide variation in the correlation coefficients for the OM for this investigation compared with investigation 1. The correlation coefficients range from 0.989 (p2 versus p6) to −0.770 (p3 versus p5).

Table 4.4

Correlation coefficients for pairs of p-values from investigation 2

	p1	p2	p3	p4	p5	p6	p7
p1	1						
p2	0.576	1					
p3	0.346	-0.552	1				
p4	0.515	0.983	-0.620	1			
p5	0.320	0.946	-0.770	0.968	1		
p6	0.607	0.989	-0.528	0.987	0.946	1	
p7	0.373	0.972	-0.728	0.977	0.988	0.955	1

The scatterplots of the p-values for this investigation reveal magnified differences between the ordering mechanisms compared with investigation 1. As for investigation 1, the ordering mechanisms which least agree on the overall value of extreme points with respect to non-extreme points are OMu3 and OMu5 ($r = -0.770$). Here there is almost a reversal of ordering since there is negative correlation. The reversal is clearly seen in the scatterplot of the p-values for these two ordering mechanisms (Figure 4.35).

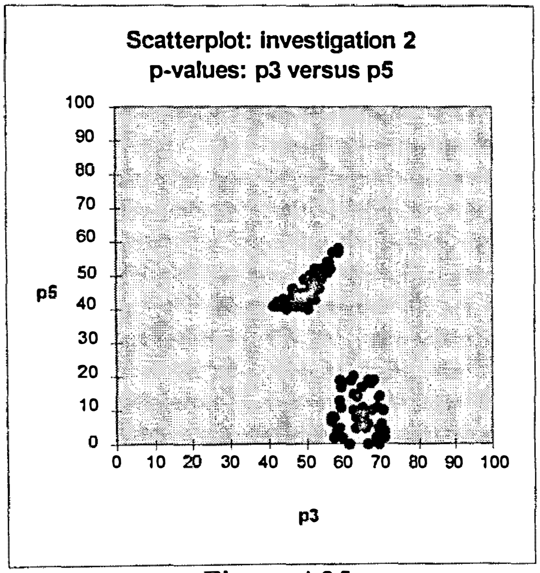


Figure 4.35

The scatterplot shows two distinct clusters of points. The cluster in the centre is aligned along the agreement line and is the p-value mapping of the non-extreme points in the assessment space.

There are differences in the p-values awarded by OMu3 and OMu5 for these non-extreme points since most of the points do not lie on the agreement line, but these differences are not great compared with the differences observed for the extreme points, which is the second cluster lying along the horizontal axis in the scatterplot. The negative correlation coefficient is due to the positioning of the second cluster below and to the right of the first cluster. This means that OMu3 is ranking the extreme points just above the non-extreme points, but OMu5 is ranking the extreme points way below the non-extreme points, hence a reversal of order for these two groups therefore negative correlation.

The greatest degree of correlation is between OMu2 and OMu6 ($r = 0.989$). An inspection of the scatterplot of p2 versus p6 (Figure 4.36) reveals perfect agreement for some points in the assessment space, evidenced by the alignment of the points along the agreement line. For other

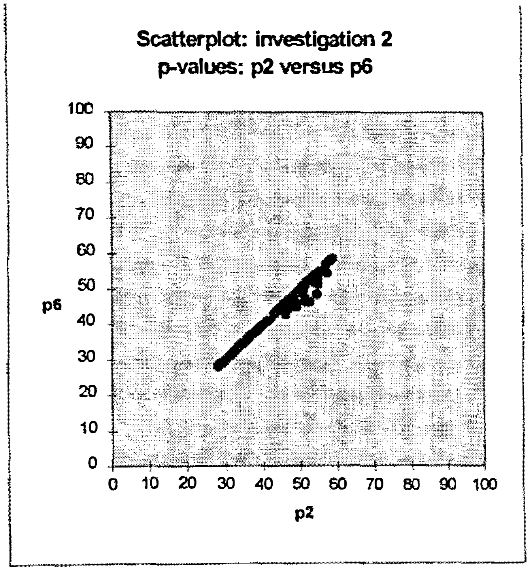


Figure 4.36

points there is close but not perfect agreement. Those points which lie along the agreement line are those points in the assessment space which lie above the target curve. The reason for this is due to the identical nature of the equivalence curves above the target curve for OMu2 and OMu6, since $m = 2$ for OMu2 and $k = 2$ for OMu6.

The points which lie below the agreement line in Figure 4.36 are the points for the non-extreme group which lie below the target curve in the assessment space. This is because the equivalence curves below the target curve for these two ordering mechanisms are different, i.e. different shapes. OMu6 is awarding lower p-values to points below the target curve than OMu2, since these points lie below the agreement line in Figure 4.36.

If the extreme points were generated in the region $80 \le x \le 100, 0 \le y \le 20$ then the correlation coefficient for these two ordering mechanisms would be much lower than 0.989 since the equivalence curves are different in the region of the assessment space below the target curve.

The remainder of the discussion for this investigation will compare OMu2 to OMu7 with OMul.

A bar chart showing the correlation coefficient between p1 and the other six OM p-values is shown in Figure 4.37. The agreement between OMul and the other six OM is much less than that for investigation 1 (Figure 4.24). This is because this investigation has generated extreme points and non-extreme points, and extreme points lie away from the target curve. Clearly, points off the target curve are awarded different p-values by each OMu (unless the equivalence curves are similar in regions of the assessment space, as is the case for OMu2 and OMu6 above the target curve). Also, the further the points from the target curve, the greater the difference in p-values.

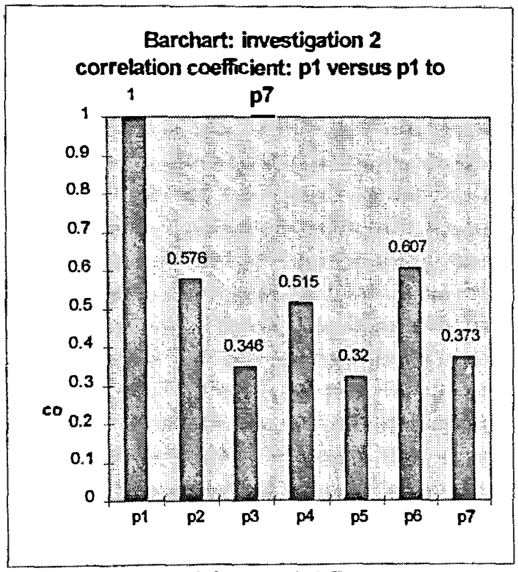


Figure 4.37

As has already been discussed, OMu2 is awarding lower p-values to points which are above the target curve, and hence to the extreme points which are above the target curve, compared with

OMu1. This is clearly seen in the scatterplot of p1 versus p2 (Figure 4.38).

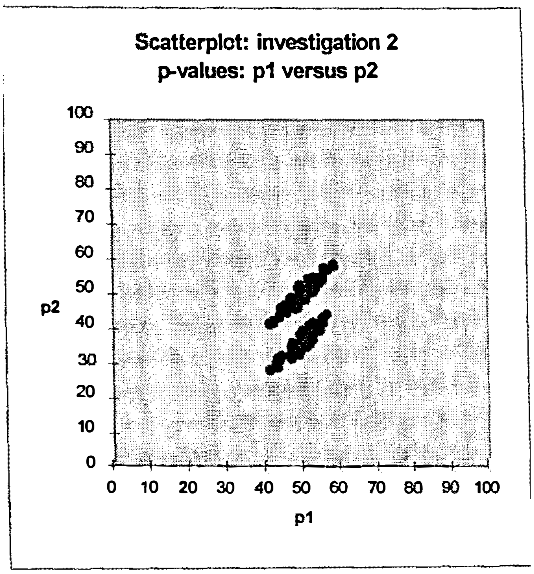


Figure 4.38

There are two clusters of points in this scatterplot. The upper cluster is the non-extreme points. This cluster is oriented along the agreement line, which is to be expected since these points lie around the target curve in the assessment space. The second cluster, which lies below the first, is for the extreme set of points. That this cluster lies directly below the first indicates that OMu2 is awarding lower p-values to the extreme points than OMu1.

In contrast with this, OMu3 is awarding higher p-values to the extreme points (Figure 4.39).

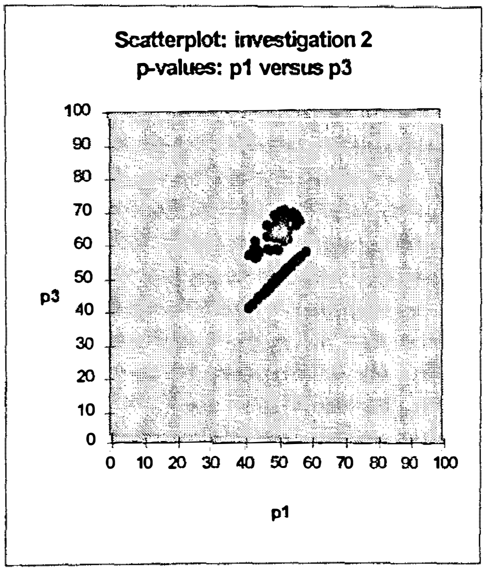


Figure 4.39

There is closer agreement between OMu1 and OMu3 for the non-extreme points since the p-values lie along the agreement line, but the second cluster which is the p-values for the extreme set of points lies above the first cluster indicating higher p-values with respect to the non-extreme group than OMu1.

OMu4, OMu5, OMu6 and OMu7 are awarding lower p-values to the extreme set of points with respect to the non-extreme set of points than OMu1, with the most noticeable difference being between OMu1 and OMu5. Scatterplots for these are shown in Figures 4.40 to 4.43 respectively.

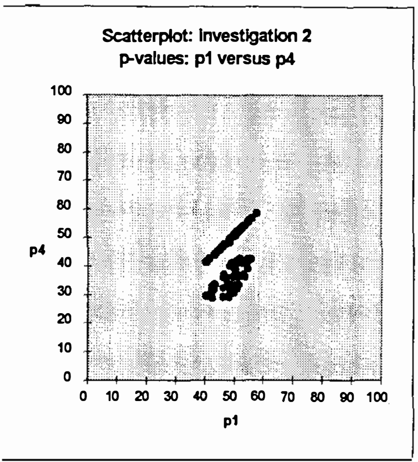


Figure 4.40

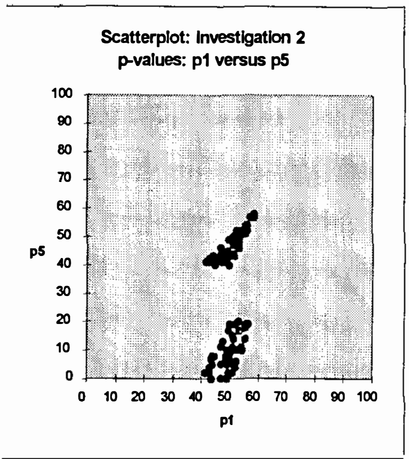


Figure 4.41

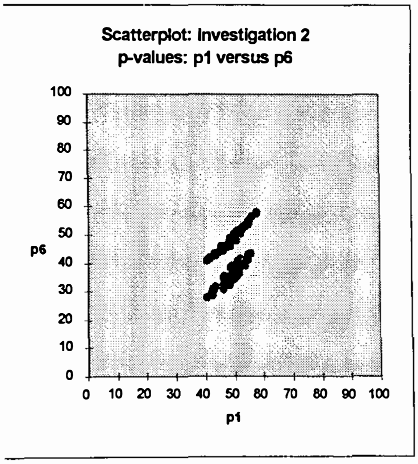


Figure 4.42

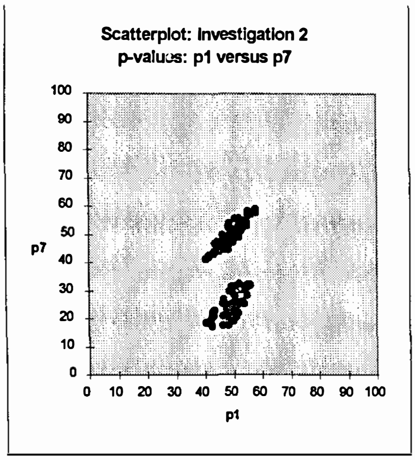


Figure 4.43

Clearly, points off the target curve are awarded different p-values for different OMu relative to points around the target curve. The choice of some OMu for an ordering system would have a considerable impact on the overall ordering of assessment data (e.g., OMu5 versus OMu1, OMu3 versus OMu7) and an impact nevertheless in other cases (e.g., OMu2 and OMu1).

This investigation has explored some of the differences between the ordering mechanisms by ordering extreme points with respect to non-extreme points in the assessment space. There was reasonable agreement between the OMu p-values for non-extreme points, and differences (considerable in some cases) between OMu p-values for extreme points.

The models developed in setion 4.3 generate different ordered distributions for points which lie off the target curve. This investigation resopnded to point 2(i) made in section 4.4.1. The following investigation will explore point 2(ii) made in section 4.4.1, i.e., the similarities between the OMu for a set of data generated around the target curve.

4.4.3 Investigation 3

Investigation 3 is a computer simulated investigation of the ordered distribution of 2-d percentage-based data for which 100 data points are generated randomly within a region confined around the target curve. The data generated in this region, to use the terminology used for investigation 2, may be called a non-extreme group of data. The program used for this investigation is in appendix F and the data generated from one run of the program is in appendix G.

A scatterplot of the assessment data generated for this investigation is shown in Figure 4.44.

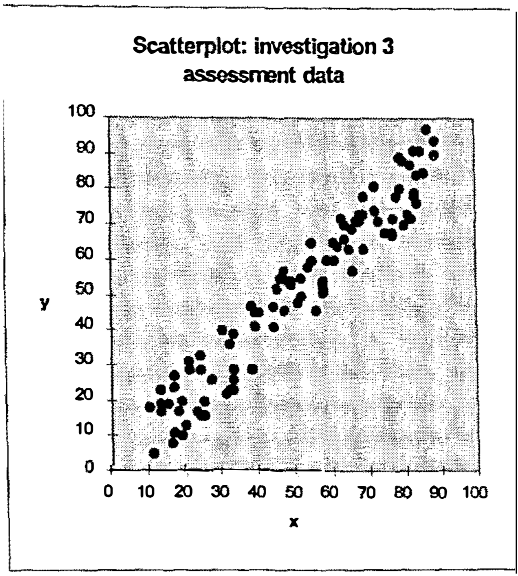


Figure 4.44

This investigation is exploring point 2(ii) made in section 4.4.1, investigation 1 that:

> Points on the target curve are awarded the same p-values by each of the seven OMu, and points in close proximity to the target curve are awarded similar p-values by each of the OMu.

This investigation will endeavour to show that points which lie around the target curve (non-extreme points) are awarded similar p-values be each of the seven OMu.

Whilst it is unlikely that a set of assessment data will lie exactly along the target curve, it is not an unrealistic scenario for authentic assessment data to lie around the target curve. Suppose x and y were related criteria. The two criteria for Senior Mathematics in Queensland (now replaced by Mathematics B and C which have a different assessment structure) are content/skill and process. Content/skill is basically the breadth of ones knowledge of facts (declarative knowledge(Biggs & Moore, 1993, p. 8)) and simple mathematical procedures (procedural knowledge (Biggs & Moore, 1993, p. 8)). Process is the ability to use that knowledge base to solve increasingly complex problems which require the synthesis and hence the deep integration (i.e., deep knowledge and understanding) of the concepts in the knowledge base. Content/skill and process are not independent criteria, even though performance in process cannot be inferred directly from performance in content/skill. (A student may memorise mathematical content and procedures and score well on a content/skill paper, but have little understanding of how concepts are related within a given problem or issue.) The ability of one to use ones knowledge base in a process question requires first that one has a knowledge base; it not unrealistic to expect that students who score well on content/skill should score better on process than students who score poorly on content/skill.

The table of correlation coefficients for this investigation is shown below (Table 4.5).

Table 4.5

Correlation coefficients for pairs of p-values from investigation 3

	p1	p2	p3	p4	p5	p6	p7
p1	1						
p2	0.998	1					
p3	0.999	0.998	1				
p4	0.999	0.998	0.999	1			
p5	0.998	0.996	0.997	0.998	1		
p6	0.999	0.998	0.999	0.999	0.999	1	
p7	0.997	0.999	0.997	0.997	0.995	0.996	1

The correlation coefficients for all pairs of OMu are greater than or equal to 0.995 indicating very close agreement. This suggests that the differences between the rank orders generated by each of the OMu for data clustered around the target curve are minimal, if not negligible for assessment purposes.

The least correlated pair of ordering mechanisms is OMu5 and OMu7. The scatterplot of their p-values is shown in Figure 4.45.

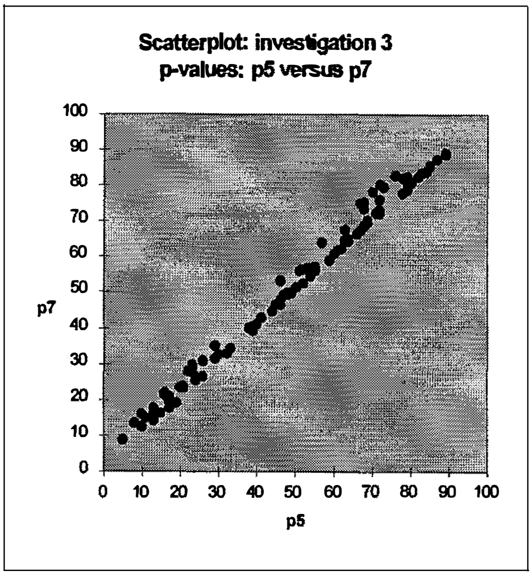


Figure 4.45

The points appear to lie on or above the agreement line, indicating that OMu7 is awarding higher p-values to points off the target curve than OMu5 (as would be expected, see previous discussions of OMu5).

The OMu which correlate highest with OMu1 are OMu3, OMu4 and OMu6 (r = 0.999). OMu3 and OMu4 generate and almost identical distribution with respect to OMu1, as seen in the scatterplots (Figure 4.46 and 4.47 respectively.

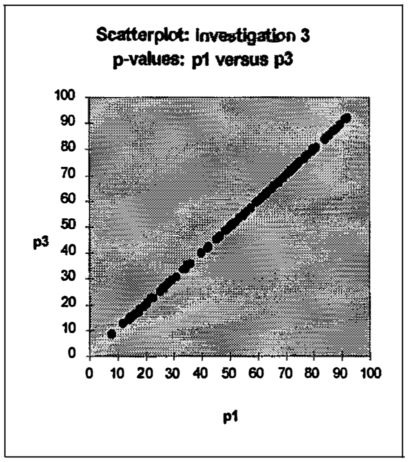


Figure 4.46

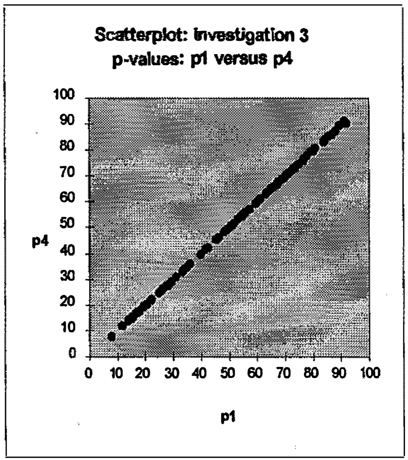


Figure 4.47

The scatterplot for OMu6 is shown in Figure 4.48.

Both OMu2 and OMu5 have a correlation coefficient of 0.998 with OMu1. Scatterplots for the p-values for these two pairs of distributions are shown in Figure 4.49 and 4.50 respectively. OMu7 is the least correlated with OMu1, r =0.997 (Figure 4.51)

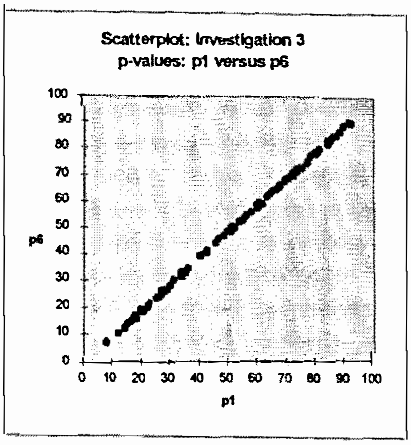


Figure 4.48

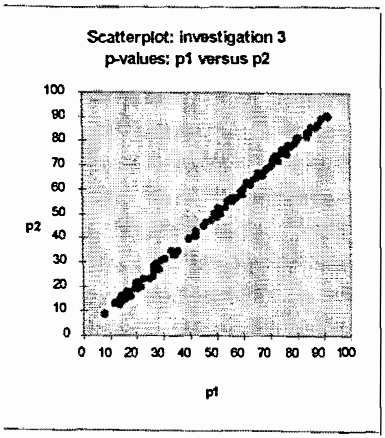


Figure 4.49

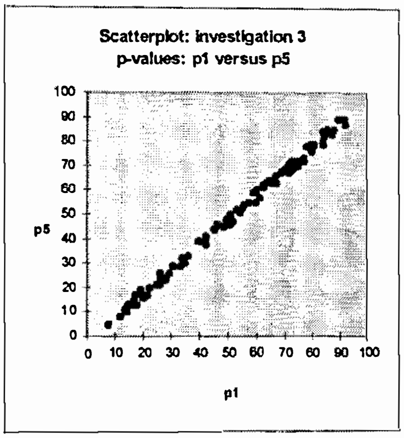


Figure 4.50

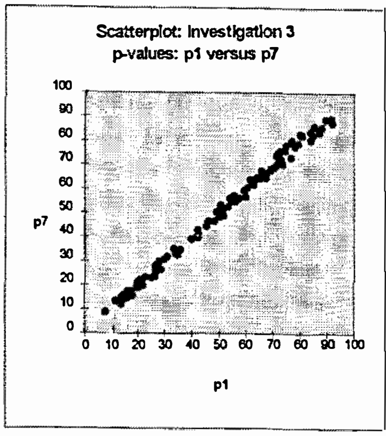


Figure 4.51

This investigation has shown that the differences between the p-values awarded to points which lie around the target curve are minimal if not negligible. This investigation supports point 2(ii) made in section 4.4.1.

The seven p-algorithms used thus far have been for 2-d assessment data. Many subjects, however, have more than 2 criteria. The following 3 investigations (4, 5 and 6) explore the properties of a set of four 3-d p-algorithms generalised from OMu1, OMu3, OMu4 and OMu5. These four p-algorithms will then be used to generate R6 and SAI distributions for actual data for a cohort of senior chemistry students (investigation 7, section 4.5).

4.4.4 Generalising 2-d p-algorithms to 3-d

The four OMu used for investigation 4, 5 and 6 are OMu1, OMu3, OMu4 and OMu5. The reason for selecting these OMu is that they do not contain extraneous constants (i.e., *m* for OMu2, *k* and *h* for OMu6, *f* and *g* for OMu7) and are therefore less variable. The 3-d algorithms for these OMu are shown below (Table 4.6). The derivation of these algorithms is not provided.

Table 4.6

3-d p-algorithms for OMu1, OMu3, OMu4 and OMu5

$p1 = \frac{x + y + z}{3}$	$p4 = 100 - \sqrt{\frac{(x - 100)^2 + (y - 100)^2 + (z - 100)^2}{3}}$
$p3 = \sqrt{\frac{x^2 + y^2 + z^2}{3}}$	$p5 = x$ if $x \le y$ and $x \le z$ $= y$ if $y \le x$ and $y \le z$ $= z$ if $z \le x$ and $z \le y$

4.4.5 Investigation 4

This is the fourth of six computer simulated investigations. In this investigation each of the four 3-d p-algorithms shown in Table 4.6 will be used to order trivariate data generated randomly throughout the 3-d assessment space. In this way, investigation 4 is similar to investigation 1, since both investigations order data which is generated throughout the entire assessment space. The program used for this investigation is in appendix H. The data for one run of the program is in appendix I.

A scatterplot of the assessment data is shown in Figure 4.52, 4.53 and 4.54. The data is 3-d hence the three projections onto the *x-y*, *x-z* and *y-z* planes.

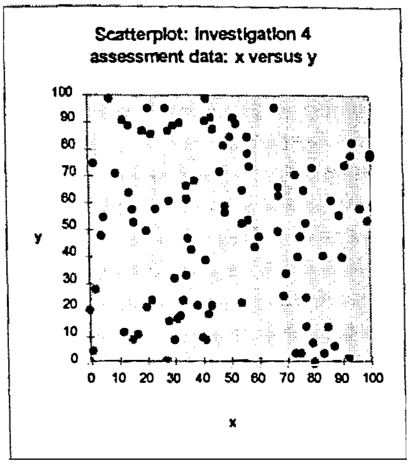


Figure 4.52

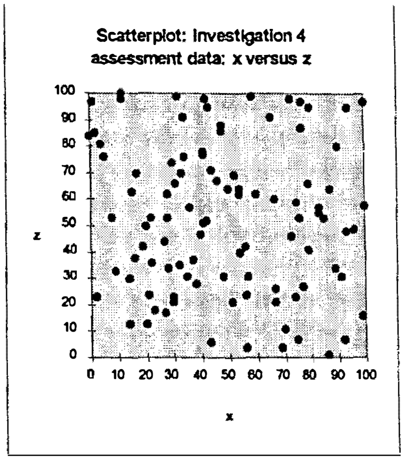


Figure 4.53

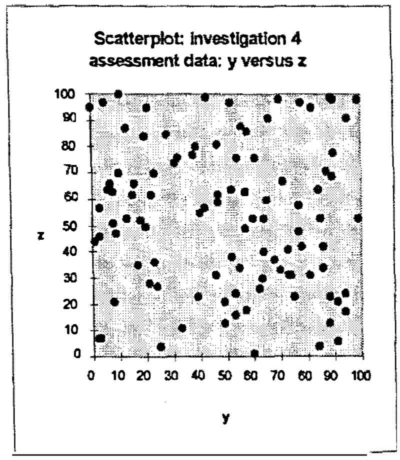


Figure 4.54

The correlation coefficients are shown below (Table 4.7).

Table 4.7

Correlation coefficients for pairs of p-values from investigation 4

	p1	p3	p4	p5
p1	1			
p3	0.954	1		
p4	0.966	0.852	1	
p5	0.792	0.616	0.904	1

The least correlated pair of OMu are OMu3 and OMu5 (r = 0.161). These are also the least correlated OMu for investigation 1 (section 4.3.1, r = 0.741). A scatterplot of p3 versus p5 is shown in Figure 4.55.

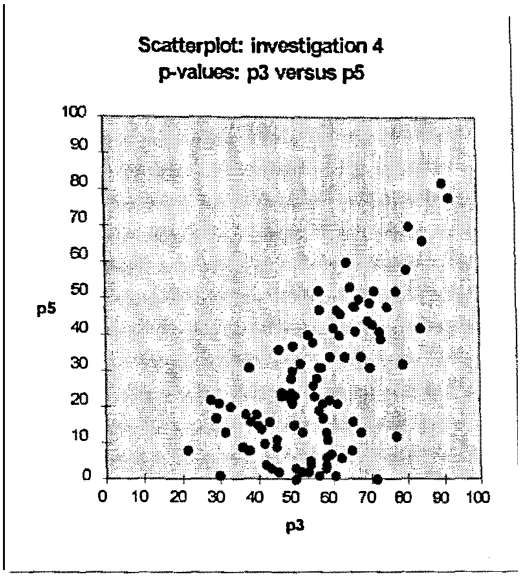


Figure 4.55

The distribution of points in the scatterplot in Figure 4.55 is similar to the p3 versus p5 distribution for investigation 1 (Figure 4.22). All of the points lie entirely below the agreement line which for this plot indicates that OMu3 is awarding higher p-values than OMu5. There are points near the agreement line but fewer than for investigation 1. This might be explained through the increased degree of freedom which the data has in 3-d assessment space with respect to 2-d assessment space. Like for the 2-d p-algorithms, the 3-d algorithms map the data onto a line (the target curve). For the 2-d p-algorithms, equivalence is modelled using equivalence curves. For the 3-d p-algorithms it can be shown that equivalence is modelled using equivalence surfaces, abstracted from the 2-d equivalence curves for the 3-d assessment space. The increased degree of freedom means a decrease in the probability that two OMu will assign the same p-value to the data due to a decrease in the probability that the assessment data will lie on the target curve. This hypothesis, however, is pursued no further in this study, except to say that it may be supported by the decrease in correlation coefficient of 6.3% (taking the range of r to from -1 to 1).

The difference between the rank orders for p3 and p5 is quite remarkable for some points in the assessment space. This is observed in the scatterplot in Figure 4.55. If the edge of a ruler is passed through the scatterplot such that the edge is parallel to the vertical axis, the points which lie along or near the edge are awarded the same (or almost the same) p-value by OMu3. The variation of points along this edge is the variation in p-values awarded by OMu5 for those points considered to be nearly equivalent by OMu3. Consider the two points from the table of data in appendix I: point 42 (52, 89, 69) and point 97 (80, 0, 95). According to OMu5, point 42 is awarded 52% overall and point 97 0%. According to OMu3, both points are awarded 72% overall.

The difference in p-values can be shown to be reflected in the determination of an SAI for these students. One OMsai involves the transformation of data from the scale of p-values to the SAI scale in such a way that the gaps between the p-values are preserved on the SAI scale (section 4.5). The SAI for any given student can be determined from their p-value and the p-values of the top and the bottom students (highest and lowest p-value respectively) in the cohort. This is shown in the following equation.

$$SAI_i = 200 + \left(\frac{p_i - p_b}{p_t - p_b} \right) \times 200$$

where SAI_i is the SAI for student i, p_i is the p-value for student i, p_t is the p-value for the top student, and p_b is the p-value for the bottom student.

The SAIs for points 42 and 97 have been determined using this formula and are shown in Table 4.8.

Table 4.8

The SAIs for points 42 and 97

OMu	SAI_{42}	SAI_{97}
3	343	343
5	333	200

Both points are awarded the same SAI from OMu3 and different SAIs from OMu5. This may seem to follow from the fact that both points are awarded lower p-values by OMu5 than by OMu3, but lower p-values do not necessarily mean lower SAIs. SAIs, as modelled by the formula, are a relative measure of achievement and therefore depends not on the magnitude of a students p-value per se but rather on the relative position of the students p-value and the p-values of the top and the bottom student (i.e. greatest and lowest p-value) in the cohort. A students SAI does not only depend on how an OMu assigns a p-value to that students results but also to the top and bottom student. This is illustrated in the SAIs for point 47. A 28% drop in p-value is reflected in only a 3% drop in SAI. It can be shown that a drop in p-value may actually result in an increase in SAI from one OMu to another.

One major impact of different OMu on MaFAD in Queensland secondary schools is how the differences in the OMu are manifest in SAI distributions. The two SAIs for point 97 calculated above show a dramatic decrease in SAI from OMu3 to OMu5 and one which would have a significant bearing on this point's Overall Position (OP). The important point is that the way the components of the multivariate data are combined to yield an overall result is a major determinant in the derivation of MaFAD from multivariate MiFAD. Point 97 is awarded 343 or 200 for the same raw MiFAD.

The scatterplots for p1 versus p3, p1 versus p4 and p1 versus p5 are shown in Figure 4.56, 4.57 and 4.58 respectively.

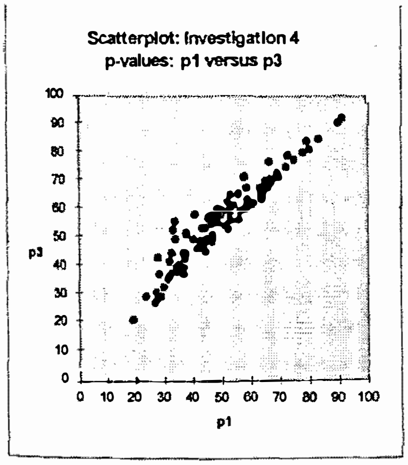


Figure 4.56

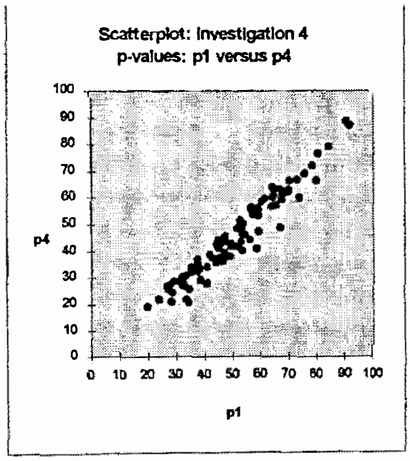


Figure 4.57

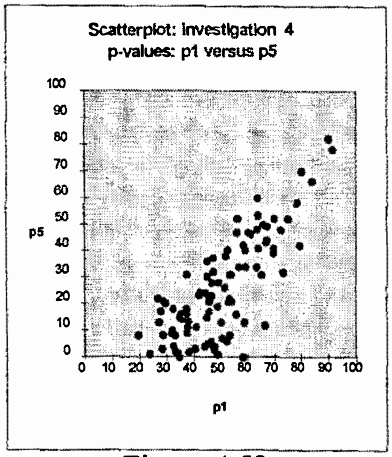


Figure 4.58

The distribution patterns for this investigation are similar to those for the corresponding scatterplots for investigation 1. OMu3 is awarding higher p-values than OMu1 (Figure 4.56) evidenced by the bulk of the points in the scatterplot lying above the agreement line (compare this with the corresponding scatterplot in investigation 1, Figure 4.29). OMu4 is awarding lower p-values than OMu1 (Figure 4.57) since the bulk of the points lie below the agreement line. This distribution is similar to that for the corresponding scatterplot in investigation 1, Figure 4.28.

This investigation has shown that the relationships between the OMu for 2-d assessment data are preserved for 3-d versions of the 2-d p-algorithms. This is shown by similar patterns of distributions of points in the scatterplots for investigation 1 and investigation 4. This suggests that an OM is not dimension specific, but rather represents general principles derived for a particular decision structure. The p-algorithms are the dimension specific manifestation of the OM.

The following investigation (investigation 5) will explore the way the four 3-d p-algorithms used in investigation 4 order assessment data generated in a non-extreme region and an extreme region of the assessment space. This investigation will show that the patterns of differences between extreme and non-extreme points for different ordering mechanisms for 2-d assessment space are preserved for 3-d assessment space. In this way, investigation 5 is a 3-d version of investigation 2.

4.4.6 Investigation 5

This is the fifth of six computer simulated investigations. In this investigation, each of the 4 3-d p-algorithms shown in Table 4.6 will be used to order trivariate data generated in two regions of the assessment space. The first region is a non-extreme region such that the points (x, y, z) are generated in the region $40 \leq x \leq 60$, $40 \leq y \leq 60$, $40 \leq z \leq 60$. The second region is an extreme region such that (x, y, z) are generated as $0 \leq x \leq 20$, $0 \leq y \leq 20$, $80 \leq z \leq 100$. The program for this investigation is in appendix J and the data generated for one run of the program is in appendix K. Scatterplots of the assessment data are shown in Figure 4.59, 4.60 and 4.61.

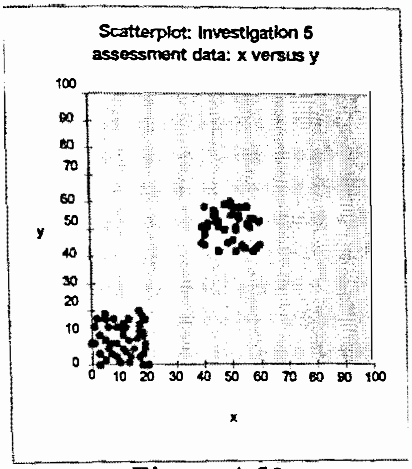


Figure 4.59

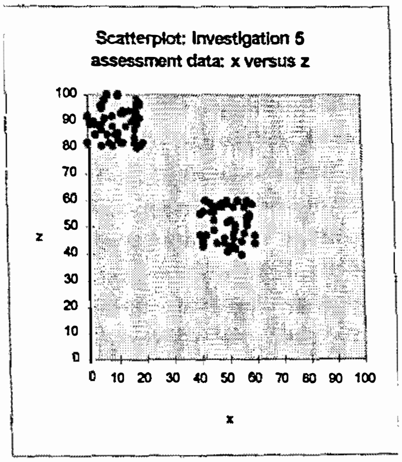


Figure 4.60

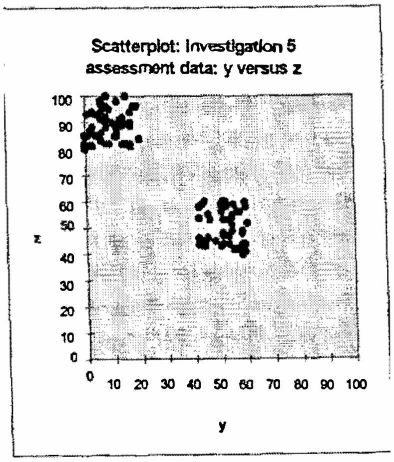


Figure 4.61

For this investigation, only three scatterplots will be discussed: p1 versus p3, p1 versus p4 and p1 versus p5.

The scatterplot of p1 versus p3 is shown in Figure 4.62. The non-extreme points have been mapped in close proximity to the agreement line which is similar to the way the non-extreme points were mapped for the 2-d investigation (see Figure 4.39). There is a difference, however, between the way OMu1 and OMu3 map the extreme points with respect to the non-extreme points between investigation 2 and this investigation. For investigation 2, p1 is awarding similar p-values to the extreme group with respect to the non-extreme group. For this investigation p1 is awarding lower p-values to the extreme group with respect to the non-extreme group since both clusters occupy approximately the same vertical positions but the extreme cluster has a lower horizontal position than the non-extreme cluster. Despite this particular difference, p3 is still favouring extreme points with respect to non extreme points, with respect to p1.

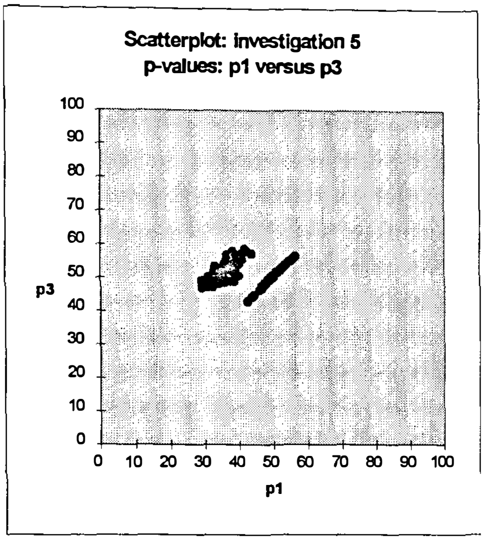


Figure 4.62

The differences between the mapping of the extreme points with respect to non-extreme points between 2-d and 3-d can possibly be explained by the increased degree of freedom for the 3-d data. For a 2-d assessment space there are two corners of the assessment space which do not lie on the target curve (Figure 4.63). These corners are (100, 0) and (0, 100) and are circled in the diagram.

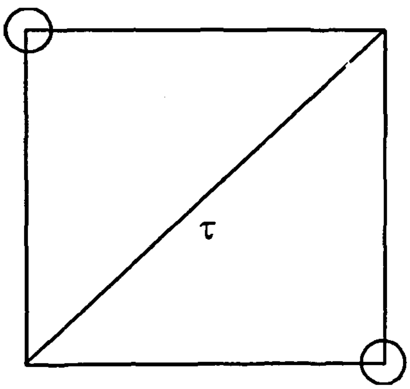


Figure 4.63: Extreme regions for the 2-d assessment space

For the 3-d assessment space there are six corners which do not lie on the target curve. These corners are shown circled in the diagram in Figure 4.64.

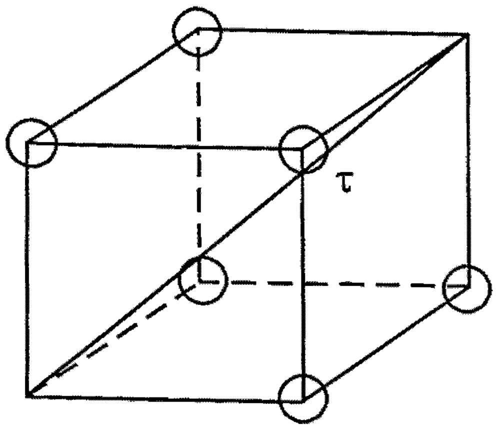


Figure 4.64: Extreme regions for the 3-d assessment space

The corners are (100, 0, 0), (0, 100, 0), (0, 0, 100), (100, 100, 0), (100, 0, 100) and (0, 100, 100). The space around these circled corners lying inside the assessment space indicate extreme regions of the assessment space. For 3-d assessment space there are two sets of extreme points. If L is a low score (say, ≤ 20) and H is a high score (say ≥80) then for the 3-d assessment space, one set of extreme points has two low scores and one high score, i.e. (L, L, H) in any order, and the other set of extreme points has one low score and two high scores, i.e. (L, H, H) in any order. For 2-d assessment space, the only possibility is (L, H) in any order. For 3-d assessment space there are two: (L, L, H) and (L, H, H). The two sets of extreme points for 3-d will be mapped differently by the OMu. The set of extreme points used here is (L, L, H). If (L, H, H) were used, different scatterplots would be obtained. This introduces new nuances into the notion of extreme data in the assessment space which could be pursued further. For example, a set of extreme points for a given n-dimensional assessment space can be identified as the number of Ls contained in the set. For the 2-d assessment space, only one possible set of extreme points exist. For assessment spaces where n>2, more than one possible set of extreme points exist, and the issue of how ordering mechanisms map different sets of extreme points with respect to non-extreme points for a given n-dimensional assessment space is a study in itself. This issue, however, will be pursued no further here.

The scatterplot of p1 versus p4 is shown in Figure 4.65. The non-extreme points are mapped around the agreement line whereas the extreme points are mapped below the agreement line. This suggests that OMu1 and OMu4, and also OMu3 discussed previously, map non-extreme points in a similar way but differences between the ordering mechanisms occur in the way the extreme points are mapped, particularly with respect to the non-extreme points. Furthermore, p4 is mapping extreme points lower the p1 and p3. Hence, OMu4 considers extreme points to be of less overall value than non-extreme points, particularly compared with OMu1 and OMu3.

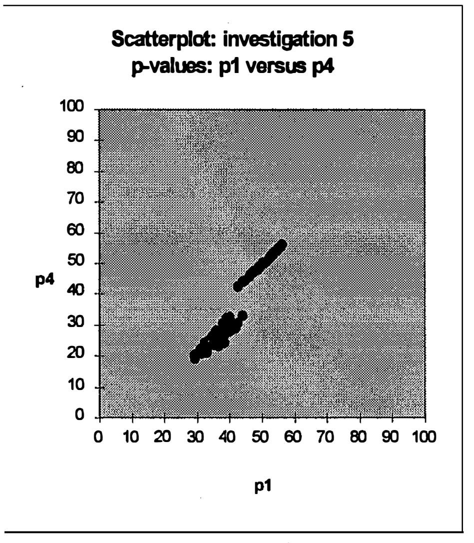


Figure 4.65

Whilst there are differences between the scatterplots for p1 versus p3 for investigation 2 and p1 versus p3 for this investigation, and differences between p1 versus p4 for investigation 2 and p1 versus p4 for this investigation, the scatterplots for p1 versus p5 for investigation 2 and for this investigation are the least dissimilar. The scatterplot for p1 versus p5 is shown in Figure 4.66. The agreement between OMu1 and OMu5 for the non-extreme points is less than the agreement between either OMu1 and OMu3, or OMu1 and OMu4. Although the non-extreme points are clustered around the agreement line, the bulk lie below the agreement line indicating OMu5 is awarding lower p-values to these points that OMu1. The extreme points, like for investigation 1, are awarded much lower p-values with respect to the non-extreme points by OMu5.

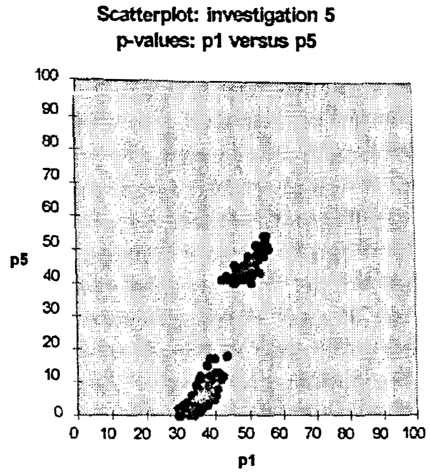


Figure 4.66

OMu5 does not distinguish between different sets of extreme points. The result (0, 0, 0) is equivalent to the result (0, 100, 100). For the 4-d assessment space, (0, 0, 0, 0) is equivalent to (0, 100, 100, 100).

Like investigation 2, this investigation has shown that differences between the ordering mechanisms are manifest, and indeed magnified, in the way the ordering mechanisms order extreme points with respect to non-extreme points. The following investigation (investigation 6) will show how the differences between the ordering mechanisms are reduced for points which lie around the target curve in a 3-d assessment space.

4.4.7 Investigation 6

This is the last computer simulation investigation. In this investigation the 4 3-d p-algorithms are used to order data generated in non-extreme regions of the assessment space. The program used for this investigation is in appendix L and the data generated for one run of the program is in appendix M.

A scatterplot of the assessment data in appendix M is shown below (Figure 4.67, 4.68 and 4.69)

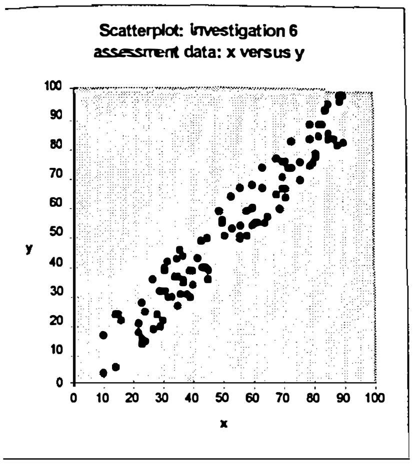


Figure 4.67

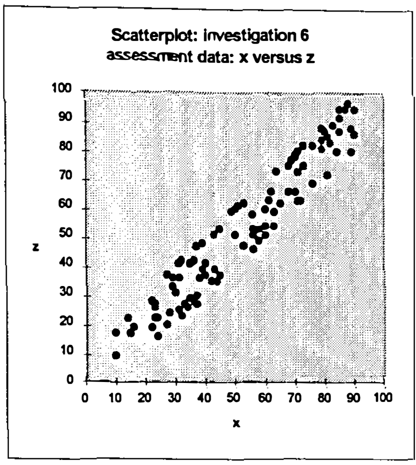


Figure 4.68

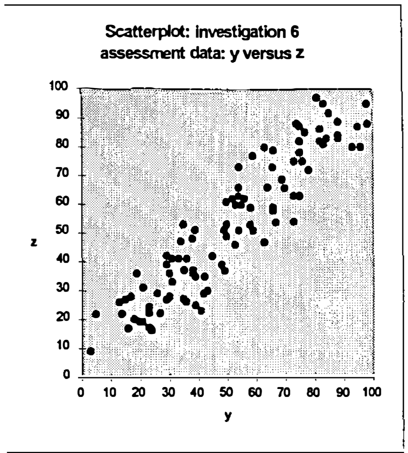


Figure 4.69

Only three scatterplots will be discussed here: p1 versus p3, p1 versus p4 and p1 versus p5.

The correlation coefficient for p1 versus p3 and p1 versus p4 is 0.999. The scatterplots, as would be expected from investigation 4.3.3, are almost identical. The scatterplots for p1 versus p3 and p1 versus p4 are shown in Figures 4.70 and 4.71 respectively. The results for non-extreme 3-d assessment data are in total agreement with the results obtained for non-extreme 2-d assessment data. The differences between the p-values awarded by OMu1, OMu3 and OMu4 for non-extreme assessment data are negligible.

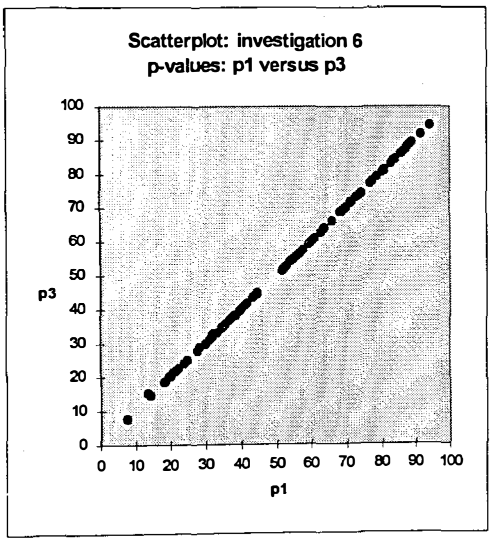


Figure 4.70

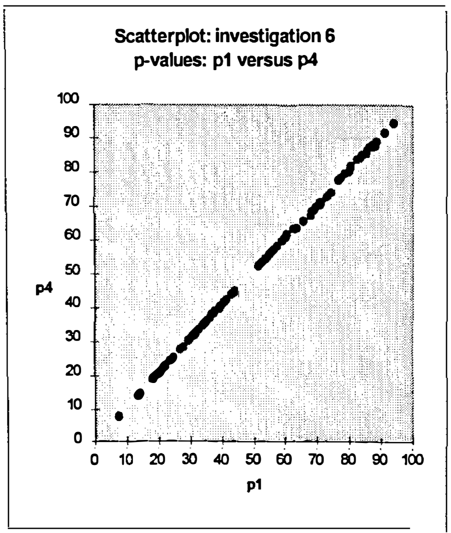


Figure 4.71

In previous investigations, OMu1 has been least correlated with OMu5. This is also the case here.

For this investigation, r = 0.996 for p1 versus p5 (Figure 4.72). The differences between OMu1 and OMu5, however, are almost negligible for non-extreme data, since this data lies in close proximity to the agreement line in the scatterplot.

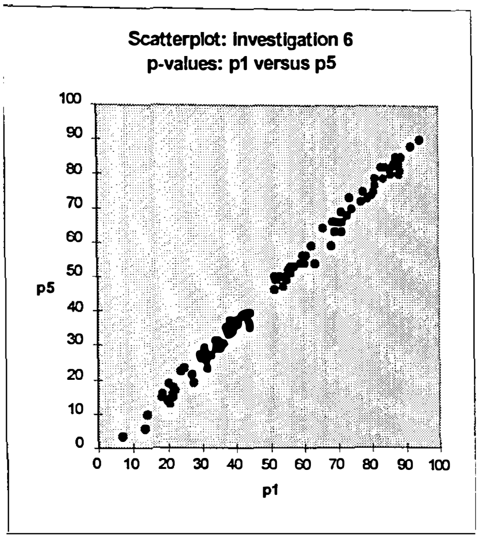


Figure 4.72

These findings are consistent with those from investigation 3, namely, that the p-values awarded by different OMu do not greatly differ for assessment data in close proximity to the target curve.

4.4.8 Findings from computer simulated investigations

The six computer simulation investigations discussed here respond to parts 1 and 2 of the research question, and furthermore, provide answers to these questions. These answers will now be addressed.

The first part of the research question is: is there a mathematical system which orders all points in a multidimensional percentage-based assessment space? The mathematical system developed in section 4.2 is such a system. The OM models developed from this system (section 4.3) have been shown to order all points generated in the assessment space. Hence, an answer to this part of the research question can be provided: there is a mathematical system which orders all points within a multidimensional percentage-based assessment space.

The second part of the research question is: is there a mathematical system which models preferences within a multidimensional percentage-based assessment space? The seven OM developed in section 4.3 were shown to generate inherently different ordered distributions of the computer simulated assessment data. These differences are manifest in the way points off the target curve are mapped to different positions on the target curve by the different OM. The two inportant findings are:

- Points off the target curve are awarded different p-values by each OMu. The further the points from the target curve, the greater the differences in the p-values awarded by each of the OMu (investigation 2 and 5).

- Points on the target curve are awarded the same p-values be each of the seven OMu, and points in close proximity to the target curve are awarded similar p-values by each of the OMu (investigations 3 and 6).

The different p-values do not simply represent a different scaling of the same inherent ordering of the data. The order for each OMu is inherently different since similar p-values are awarded to points around the target curve but different p-values to points off the target curve.

The differences between the distributions for the set of OMu studied here represent different preferences for regions in the assessment space. For example, consider a set of points which are equivalent with according to OMu1. For this set, OMu5 prefers points on and around the target curve over points off the target curve. OMu3 prefers points off the target curve over points around the target curve. OMu4, like OMu5, prefers points around the target curve, but maps points off the parget curve higher that OMu5. OMu2 prefers points on one side of the target curve over points on the other side of the target curve.

These findings peovide an answer to the second part of the research question: there is a mathematical system which models preferences within a multidimensional percentage-based assessment space.

The third part of the research question calls for the design of an OS using the mathematical system developed in section 4.2. This is addressed by investigation 7 in section 4.5.

4.5 Investigation 7

This investigation applies the 4 3-d p-algorithms used for investigations 4, 5 and 6 to a sample of CBAD for a cohort of senior chemistry students from a Brisbane secondary school. This investigation seeks to show how the OMu developed in section 4.3, and hence the mathematical system developed in section 4.2, can be used to construct an independent ordering system for a percentage-based assessment system (research question part 3, section 1.5) to generate R6 and SAI MaFAD from CBAD MiFAD.

4.5.1 The ordering system used for investigation 7

The assessment system for this cohort of students generates trivariate percentage-based assessment data. The final assessment data was obtained for 54 students, together with the R6 and SAI distributions generated by the science department using this CBAD. Of these 54 students, 6 were not used for this investigation as these students did not complete 4 semesters of study. Hence 48 students from this cohort were used.

The three criteria defining achievement in senior chemistry are: content (x), applied process (y) and scientific process (z). The decision table used by the school to define LOAs in terms of minimum cutoffs in each criterion is shown in Table 4.9.

Table 4.9

The decision table used by the school to determine LOA.

LOA	x	y	z
VHA	≥ 80	≥80	≥ 80
HA	≥ 65	≥ 65	≥ 65
SA	≥ 50	≥ 45	≥ 45
LA	≥ 30	≥ 30	≥ 30
VLA	< 30	< 30	< 30

Apart from the decision table, no other information was supplied pertaining to the OS used by the school to generate the R6 and SAI distributions.

The decision table shown above cannot be used by OMu5 without modification of the p-algorithm. This is because the cutoffs for each LOA must lie on the target curve due to the way the OMu5 was designed in section 4.2.1. The modification of the p-algorithm would involve re-defining the target curve and equivalence curves in the assessment space. The nature of this redefinition can be explained using the 2-d assessment space. If the cutoffs for VHA in 2-d was (80, 80) and HA was (70, 60), the following set of equivalence curves and target curves would arise (Figure 4.73)

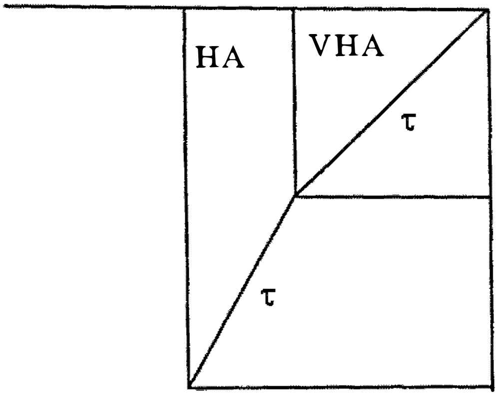


Figure 4.73: OMu5 for HA and VHA cutoffs for school's decision table

This model could be used to generate a p-algorithm (not done for this investigation) in a similar way to that used in section 4.2.1, but the p-algorithms would be slightly more complicated than that for OMu5. The restriction that $x = y = z$ for the LOA cutoffs is only a feature of OMu5 and OMu6, not the other 5 OM.

In light of the restriction placed on the cutoffs by OMu5, the decision table has been modified slightly to allow for the use of OMu5 in this investigation. The criterion x cutoff for SA has been changed from 50 to 45. The decision table used for this investigation is shown in the following table (Table 4.10).

Table 4.10

The cutoffs used for investigation 7

LOA	x	y	z
VHA	≥ 80	≥80	≥ 80
HA	≥ 65	≥ 65	≥ 65
SA	≥ 45	≥ 45	≥ 45
LA	≥ 30	≥ 30	≥ 30
VLA	< 30	< 30	< 30

The computer program used to generate the R6 and SAI distributions is in appendix N. The data provided by the school, together with the MaFAD generated by the program is in appendix O. The OS used by the program will now be explained.

The general structure of the independent OS (Figure 1.16 section 1.3) has been reproduced below.

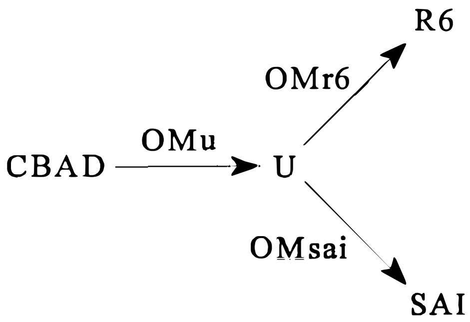


Figure 4.74: Independent ordering system

The 3-d p-algorithms for OMu1, OMu3 OMu4 and OMu5 are each used to transfer CBAD onto U. Hence, 4 distributions are generated, one for each OM.

The OMr6 and OMsai are the same for each of the four OMu. OMr6 depends on the cutoffs in the decision table. A CBAD point is positioned on the R6 in the following manner (this procedure is OMr6).

- The LOA cutoffs are mapped onto U using each of the OM. (Since the cutoffs lie on the target curve, p-value LOA $= x = y = z$, i.e., the scores which define minimum achievement in the decision table for each LOA).
- The interval along U defined by the minimum p-value for a LOA and the minimum for the next highest LOA is subdivided into 10 intervals of equal length. For VHA the next highest p-value is 100. For VLA, the lower p-value is 0.
- CBAD values are awarded a rung on the R6 depending on the band and rung interval into which the corresponding p-value falls on the U scale.

OMsai is a transformation for the p-values on U such that

$$SAI_i = 200 + \left(\frac{p_i - p_b}{p_t - p_b} \right) \times 200$$

where SAI_i is the SAI for student i, p_i is the p-value for student i, p_t is the p-value for the top student, and p_b is the p-value for the bottom student (section 4.4.5).

Other OMsai can be identified. For example, suppose the teachers in a department believed that it is more difficult for a student to move through the VHA LOA than the HA LOA. That is, the movement from one rung to the next highest rung represents more in overall achievement for a VHA student than a HA student. This element of a decision structure could be easily structured into an OMsai.

If the data were mapped such that different LOA bands had different scalings applied to the transformation of the data onto the SAI scale, then an intermediatory step could be included as a part of OMsai. This intermediatory step would involve the transformation of data from U to another univariate scale such that the transformation involves a stretching of some LOA bands and a compression of others. If the VHA band is considered 1.5 times more 'difficult' to move through than the HA band, then the VHA region on U would be stretched 1.5 times with respect to the HA region on U.

It should be noted that such distortions of U shift the mean and standard deviation of the SAI data. This, inturn, impacts on the OP calculations. It would be possible to show that compressing some bands and stretching others may increase the chance of the VHA students obtaining higher OPs than otherwise. This issue is not explored here but could be cause for concern since schools are at liberty to apply different decision structures to the calculation of SAIs.

4.5.2 The R6 distribution

The data obtained from the school is shown in three scatterplots: x versus y (Figure 4.75), x versus z (Figure 4.76) and y versus z (Figure 4.77) From the plots of x versus y and x versus z it is clear that as a group, students score higher in x than in either y or z. The bulk of the data lie below the agreement line in both plots. This suggests that marks are more difficult to obtain in either y or z than in x. y and z are the criterion 'applied process' and 'scientific process' respectively, and x is 'content'. Process criteria are generally more challenging for students than content because process tasks generally require a deeper level of knowledge integration and understanding than content tasks.

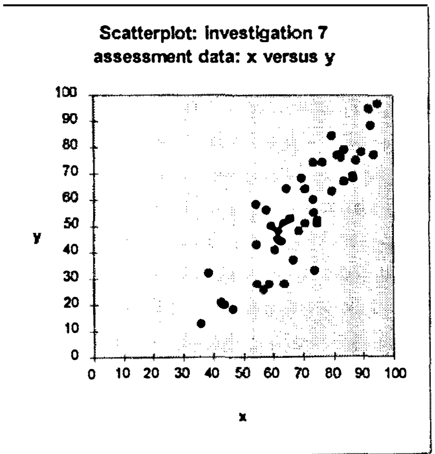


Figure 4.75

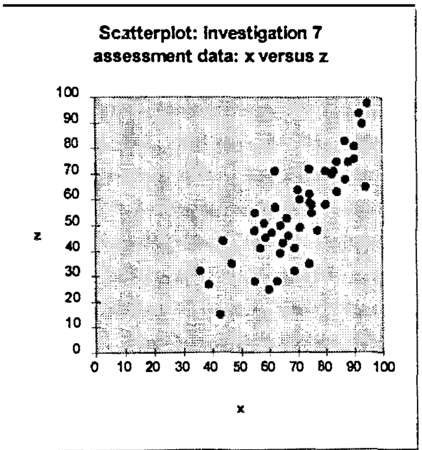


Figure 4.76

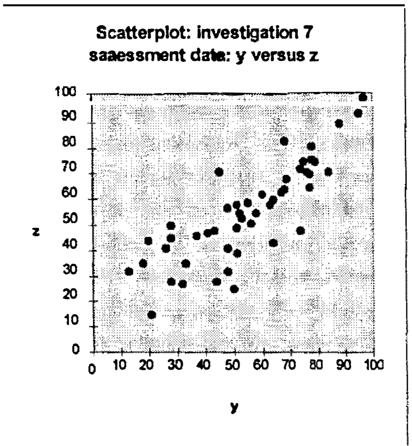


Figure 4.77

One of the findings from the computer simulation investigations in section 4.4 is that data which lies close to the target curve (non-extreme data) receive similar p-values from each of the OMu and data which lies away from the target curve (extreme data) receive different p-values. For a set of OMu which share a common target curve, the proximity of the data to the target curve is indicative of the susceptibility of the data to be awarded different p-values by each of the OMu. This can be expressed another way. The further the data is from the target curve, the more sensitive the data is to shifts in the ordering effected by different OMu. Some results in the cohort are more sensitive than others and this is reflected in the R6 distributions.

The R6 distributions are shown in Table 4.11. The term 'R6-school' is the school's R6 and R6-1, R6-3, R6-4 and R6-5 are the R6 distributions for the four OMu. The term OM-R6s refers to the four OM R6s collectively.

Students 12, 27 and 21 are the top three VHA students. The results for these students are: 12 (95, 97, 98), 27 (92, 95, 94) and 21 (93, 88, 90). For each of these students, the results in each criterion are similar; the maximum difference is between *y* and *z* for student 21 and is only 5%. These students lie very close to the target curve (i.e., in a non-extreme region of the assessment space), and should not be sensitive to different OMu. Indeed, this is observed in the R6 distribution. Student 12 is awarded a VHA9 by three of the four OMu and VHA8 by the other (OMu5). The school has also placed student 12 at VHA9. Student 27 is placed at VHA7 on the four OM-R6s and at VHA8 by the school; student 21 placed at VHA6 by the school and three OMu and at VHA5 by OMu5. Students who do well in all criteria are not sensitive to shifts in the order brought about by a different OMu because the results of these students lie in a non-extreme region of the assessment space. Similarly, students who do equally poorly in each criteria, e.g. students 13 and 16 in the VLA band (Table 4.11) are not sensitive. A student with the result (100, 100, 100) will always be a VHA10 and will always get an SAI of 400; a student with the result (0, 0, 0) will always be LA1 and get an SAI of 200.

Only one student was found to be totally insensitive to the different OMu and this is student 5. Student 5 is an SA6 on all 5 R6 distributions. Not surprisingly, student 5's results are (55, 58, 55) which are very near the target curve.

Some students received very different placements on each of the R6 distributions. For example, the placement of student 2, with the results (59, 28, 45), range from VLA to SA. The school awards student 2 an LA7, probably due to the low result in criterion *y*. This student obtained 28 and the cutoff in this criterion is 45. Both OMu1 and OMu4 award an LA, LA10 and LA9 respectively, whereas OMu3 awards an SA1. The result from OMu3 for student 2 is consistent with the findings from investigation 1 to 6 for which OMu3 was shown to favour a strength in one criteria over a weakness in another with respect to the other OMu. The result of 59 in criterion *x* is more than compensated by OMu3 for the result of 29 in criterion *y*. On the other hand, OMu5 awards student 2 a VLA 19, a result based on the 28 in criterion *z*. OMu5, the no-tradeoff OMu, disregards the students' achievement in either criterion *x* or *y*, and grades the student using the lowest component.

The issue of generating MaFAD from MiFAD and the complexities surrounding this process is evident for student 2 and others in the cohort (e.g., 45, 46, 11, 26). These students receive very different positions on the R6 using the same cutoffs in each LOA but different OMu. The nature of the combination of the data to generate a univariate result is clearly important for ordering students who fall away from the target curve.

A detailed analysis of the patterns of similarities and differences between the R6-school and the R6-OMs is beyond this study. However, an overview of the similarities and differences is provided in Table 4.12. This table presents the shift in placement for each student for each OMu with respect to the student's placement on the R6-school.

A summary of the transitions between bands shown in Table 4.12 is in Table 4.13. There are some significant differences between the LOAs awarded to the students in the cohort by the school and by each of the OMu. The most striking difference is for OMu5 which shifts 22 students down one rung, which is consistent with the properties of OMu5 identified in section 4.3 and 4.4. Of the four OMu, OMu4 yields the most similar R6 to R6-school. A total of 9 students shifted a band, 5 up one and 4 down one. The mean and standard deviation of the shifts also shows R6-4 to be the most similar to R6-School. The mean shift for OMu4 is 0.42 rungs with a standard deviation of 1.70 rungs. The mean and standard deviation (respectively) for the other OMu are Mu1: 1.08, 1.96; OMu3: 1.50, 2.10; OMu5: -3.96, 2.77.

Table 4.11: R6 distributions for investigation 7

LOA	Rung	School	OMu1	OMu3	OMu4	OMu5
	50					
	49	12	12	12	12	
VHA	48	27				12
	47		27	27	27	27
	46	21	21	21	21	
	45					21
	44					
	43	33,39				
	42	3,34	39	39	39	
	41	24	34	34	34	
	40	23,47	3,8,24,37	3,8,24,33,37	3	
	39		33		24,33,37	39
HA	38	37	23,47	23,47	8,23,47	34
	37	8,36	4	4		3,24
	36		36	36	4,36	
	35	4	25	25		33,36,47
	34				25	23
	33					4,37
	32	25,30	30,45	18,30,45	30	
	31	18	9,15,18	9,15	45	8
	30	9		6, 46	9,15,18	25,30
	29	45	19,46	19	46	
SA	28	15	6,40	40	6,19	9,15
	27		14,41,44	14,41,44	40	45
	26	5	5,28,29	5,28,29	5,14,29,41,44	5,46
	25	6,14,19,28,44,46,40		32	28	14
	24	29,41	17,32	7,17,43	32	6,19,28
	23		7,22,43	11,22,38	7,17	41
	22	17,32	11,31,38	1,26,31	22,31,43	18,29
	21	7,22,31	1,26	2	11	40
	20	1,43	2		38	
	19			35	1,2,26	31,44
LA	18	11,26,38	35			22,32
	17	2,35		10	35	17
	16			48		
	15	42,48	10,48			7
	14	10		20	10,48	
	13		20	42		38
	12	20	42		20,42	43
	11					
	10	13,16	16	13,16		2,10,11,26,42
	9		13		13, 16	1,35
VLA	8					
	7					20, 48
	6					13
	5					16
	4					
	3					
	2					
	1					

Table 4.12
Rung differences between R6-school and OM-R6s for each student
(*indicates inter LOA transition)

	OMu			
No.	1	3	4	5
1	1*	2*	-1	-11*
2	3	4*	2	-7*
3	-2*	-2*	-2*	-5*
4	2	2	1	-2
5	0	0	0	0
6	3	5	3	-1
7	2	3	2	-6*
8	3	3	1	-6
9	1*	1*	0	-2
10	1	3	0	-4*
11	4*	5*	3*	-9*
12	0	0	0	-1
13	-1	0	-1	-4
14	2	2	1	0
15	3*	3*	2*	0
16	0	0	-1	-5
17	2	2	1	-5*
18	0	1	-1*	-9*
19	4	4	3	-1
20	1	2	0	-5
21	0	0	0	-1
22	2	2	1	-3
23	-2	-2	-2	-4
24	-1*	-1*	-2*	-4*
25	3	3	2	-2*
26	3	4	1	-8*
27	-1	-1	-1	-1
28	1	1	0	-1
29	2	2	2	-2
30	0	0	0	-2
31	1	1	1	-2*
32	2	3	2	-4*
33	-4*	-3*	-4*	-8*
34	-1	-1	-1	-4*
35	1	2	0	-8*
36	-1	-1	-1	-2
37	2	2	1	-5
38	4*	5*	2	-5
39	-1	-1	-1	-4*
40	3	3	2	-4
41	3	3	2	-1
42	-3	-2	-3	-5*
43	3*	4*	2*	-8*
44	2	2	1	-6*
45	3*	3*	2*	-2*
46	4	5	4	1
47	-2	-2	-2	-5
48	0	1	-1	-8*

Table 4.13

Number of students who were shifted an LOA band by each OMu with respect to R6-school

	OMu			
	1	3	4	5
Up 1 LOA band	8	9	5	0
Down 1 LOA band	3	3	4	22
Total band shifts	11	12	9	22

One final point of interest is the cluster of students on SA5 of R6-School. Students 6, 14, 19, 28, 44, 46 and 40 are considered to be equivalent according to the school's OS. Each of the four OMu, however, discriminate between these students on the R6. It is uncertain why the school has decided to place these students in this way.

4.5.3 The SAI distribution

The SAI data obtained from the school, together with the SAI distributions calculated by each of the OMu, are in appendix O.

The correlation coefficients for investigation 7 are shown below (Table 4.14). The SAI distributions for each of OMu1, OMu3, OMu4 and OMu5 are labeled SAI1, SAI3, SAI4 and SAI5 respectively. SAI is the school's distribution.

Table 4.14

Correlation coefficients for pairs of SAIs from investigation 7

	SAI1	SAI3	SAI4	SAI5	SAI
SAI1	1				
SAI3	0.999	1			
SAI4	0.999	0.997	1		
SAI5	0.978	0.972	0.984	1	
SAI	0.981	0.979	0.984	0.972	1

The highest correlation with the school's distribution is for OMu4 (r = 0.984) whereas OMu5 correlates the least (r = 0.972).

The top three VHA students, students 12, 27 and 21, are awarded almost the same SAI by the school and each of the OMu. Student 12 is 400 on all distributions; student 27 is 393 on the school distribution and ranges from 391 to 393 on the OM distributions; student 21 is 379 on school distribution and ranges from 381 to 383 on OM distributions. This is consistent with similar placements of these students on the R6 distributions and for the same reasons outlined in section 4.5.2.

The OS developed and used here to generate the R6 and SAI distributions is an independent OS, that is, the R6 and SAIs are generated from a common univariate distribution. An inspection of the school's SAI distribution suggests that the school is using a dependent OS (Figure 1.5 (a)). Consider the seven students placed by the school on SA5 of the R6. Each of these students receives an SAI of 271 from the school. The SAI given by the school for students on the same rung of the R6 are the same, suggesting a direct mapping of students from the R6 scale onto the SAI scale. The SAIs for these six students are quite different for the OM SAI distributions: student 28 and 19, for example, are equivalent on the school's SAI distribution but differ by 18 SAI points according to OM1.

Further analysis of the SAI distributions are beyond this study. What this section has demonstrated, however, is the use of the mathematical models developed in section 4.3, and hence the mathematical system developed in section 4.2, to construct an independent OS for a percentage-based assessment system to yield SAI (and R6, in section 4.5.2) distributions. It is clear that such a system can be used to construct and independent OS for a percentage-based assessment system.

The investigation of the suitability of such an OS for school-based assessment systems, and the specific nature of the decision structures at the heart of OS used within schools, are not intended to be addressed here, but do provide possible avenues for further investigation.

4.6 Summary

This study has addressed the following research question:

RESEARCH QUESTION:

Is there a mathematical system which:

(1) orders all points in a multidimensional percentage-based assessment space,

(2) models preferences within a multidimensional percentage-based assessment space,

(3) can be used to construct an independent ordering system for a percentage-based assessment system?

The research conducted here and discussed in this chapter suggests there is such a mathematical system.

Using the concept of an indifference curve (Keeny & Raffia, 1976) this author developed a mathematical system (section 4.2) to order all points in a multidimensional percentage-based assessment space. The system consists of equivalence curves/surfaces (i.e., indifference curves) and target curves, and from these two elements are derived distance algorithms and percentage algorithms (p-algorithms). The p-algorithms are essentially value functions (Keeny & Raffia, 1976).

This author used the mathematical system in section 4.3 to develop seven ordering mechanisms, each sharing a simple linear target curve but having different shaped equivalence curves. Whilst the seven ordering mechanisms are only a sample of the many possible ordering mechanisms which could be developed, they represent a broad range of preferences and provide sufficient variety for the purposes of this study. The p-algorithms (value functions) were derived from the geometry of the curves for each ordering mechanism (Table 4.1).

The seven 2-d p-algorithms for the ordering mechanisms developed in section 4.3 were studied in three computer simulation investigations: investigation 1, 2 and 3. Two findings emerged from these investigations, the first in respect of the first part of the research question, the second in respect of the second part of the research question.

The first finding is:

1 The seven OMu developed in section 4.3 from the mathematical system developed in section 4.2, generate a p-value, and hence a position on an ordered scale, for every point generated in the assessment space. It can be shown that the seven OMu order all points in the assessment space.

The second finding is:

2 The seven OMu generate inherently different rank order distributions for data generated randomly throughout the assessment space. These differences occur in the way each of the OMu assign a p-value to points off the target curve with respect to points lying on the target curve. That is:

(i) Points off the target curve are awarded different p-values by each OMu. The further the points from the target curve, the greater the difference in p-values awarded by each of the OMu.

(ii) Points on the target curve are awarded the same p-values by each of the seven OMu, and points in close proximity to the target curve are awarded similar p-values by each of the OMu.

The differences in the ordered distributions generated by each OMu can be interpreted as differences in preference for regions in the assessment space.

Four of the seven ordering mechanisms were generalised to yield 3-d versions of the 2-d algorithms used in investigation 1, 2 and 3 (Table 4.6). These four 3-d p-algorithms were used to reproduce the first 3 investigations for 3-d assessment space (investigations 4, 5 and 6). It was found that the findings for the 2-d algorithms held for the 3-d algorithms.

The final investigation, investigation 7, achieved two things. First, an independent ordering system was developed to generate R6 and SAI distributions from percentage-based CBAD. Second, this ordering system was used to produce R6 and SAI distributions for a cohort of senior chemistry students using the four 3-d p-algorithms from investigations 4, 5 and 6. It was found that both R6 and SAI distributions could be generated from the CBAD and that the R6 and SAI distributions for the cohort for each of the 4 OMu were different, reflecting different preferences and decision structures. Also, similarities and differences between the R6 and SAIs for the OMu and the school's R6 and SAI distributions were identified. Some of the OMu distributions (e.g. OMu4) were not too different from the school's, perhaps suggesting the possibility of using either this OMu, or modifying it slightly, to generate R6 and SAIs in the department's assessment system. Perhaps this chemistry department may be able to use an independent OS with an appropriate OMu to calculate its R6 and SAI distributions. It seems, however, that the school uses a dependent OS, although the nature of this particular system is not documented. This investigation has used the mathematical system developed in section 4.2 to design and apply an independent ordering system for a percentage-based assessment system.

CHAPTER FIVE

SYNTHESIS AND REVIEW

5.1 Purpose of research

The purpose of this research is to explore the issue of generating macro-function assessment data from micro-function assessment data. More specifically, this research considers the issue of ordering multivariate educational assessment data for percentage-based assessment systems, and in particular, the nature of the ordering mechanisms for ordering systems within such assessment systems.

This study has been focused by the following research question:

RESEARCH QUESTION:

Is there a mathematical system which:

(1) orders all points in a multidimensional percentage-based assessment space,

(2) models preferences within a multidimensional percentage-based assessment space,

(3) can be used to construct an independent ordering system for a percentage-based assessment system?

5.2 Design of the study

The design of the study followed the principles of computer simulation experiments which involve the creation of a mathematical model and the exploration of the operation of that model using computer simulation.

The research for this study is divided into four sections. The first section (section 4.2) is the first of two modelling sections. It involves the creation of a mathematical system to order points in a multidimensional percentage-based assessment space. The mathematical system uses the concept of an equivalence curve (indifference curve Keeny & Raffia, 1976) and a target curve to define order in the space.

The second section (section 4.3) is the second modelling section of the research. It involves the development of specific models using the mathematical system developed in section 4.2. This section develops seven models, called OMu, and from each model is derived a value function called a p-algorithm.

The third section (section 4.4) conducts six computer simulation investigations of the models (OMu) developed in section 4.3.

The fourth section (section 4.5) applies four 3-d p-algorithms to percentage-based CBAD obtained from the chemistry department of a Queensland secondary school for a cohort of students.

5.3 Significance of the study

The emphasis on criteria, standards and profiles of achievement in Queensland education, the acknowledgement of the destructive effects which the application of mathematical models (i.e. normal distributions) have on the curriculum, and the preference of multidimensional profiles over unidimensional grades and scores, seems to have left unaddressed the important issue of how to order multidimensional data, and the necessary unidimensional nature of ordered distributions. Furthermore, the requirement that the process of determining SAIs be decisions not calculations (Viviani, 1990, p. 133) raises important questions about the use of mathematical models for assessment processes. Do 'calculations' necessarily undermine quality decisions, or can decisions be modelled using mathematics and be calculations? This study addresses this issue and seeks to articulate the structure of ordering systems, and in particular, mathematical systems which can be used to design OMu for ordering systems.

5.4 The research questions answered

Using the concept of an indifference curve (Keeny & Raffia, 1976) a mathematical system was developed to order all points in a multidimensional percentage-based assessment space. The system consists of equivalence curves/surfaces (i.e., indifference curves) and target curves, and from these two elements are derived distance algorithms and percentage algorithms (p-algorithms). The p-algorithms are essentially value functions (Keeny & Raffia, 1976).

The mathematical system was used in section 4.3 to develop seven ordering mechanisms, each sharing a simple linear target curve but having different shaped equivalence curves. p-algorithms (value functions) were derived from the geometry of the curves for each ordering mechanism (table 4.1).

The seven 2-d p-algorithms for the ordering mechanisms developed in section 4.3 were studied in three computer simulation investigations: investigation 1, 2 and 3. Two findings emerged from these investigations, the first in respect of the first part of the research question, the second in respect of the second part of the research question.

The first part of the research question is as presented in section 1.5 is: Is there a mathematical system which orders all points in a multidimensional percentage-based assessment space?

The first finding is that the seven OMu developed in section 4.3 from the mathematical system developed in section 4.2, generate a p-value, and hence a position on an ordered scale, for every point generated in the assessment space. It can be shown that the seven OMu order all points in the assessment space. Thus it has been shown that there is a mathematical system which orders all points in a multidimensional percentage-based assessment space.

The second part of the research question as presented in section 1.5 is: Is there a mathematical system which models preferences within a multidimensional percentage-based assessment space?

The second finding is that the seven OMu generate inherently different rank order distributions for data generated randomly throughout the assessment space. These differences occur in the way each of the OMu assign a p-value to points off the target curve with respect to points lying on the target curve. Two observations are made. Firstly, points off the target curve are awarded different p-values by each OMu. The further the points from the target curve, the greater the difference in p-values awarded by each of the OMu. Secondly, points on the target curve are awarded the same p-values by each of the seven OMu, and points in close proximity to the target curve are awarded similar p-values by each of the OMu.

The differences in the ordered distributions generated by each OMu can be interpreted as differences in preference for regions in the assessment space. Thus it has been shown that there is a mathematical system which models preferences within a multidimensional percentage-based assessment space.

Four of the seven ordering mechanisms were generalised to yield 3-d versions of the 2-d algorithms used in investigation 1, 2 and 3 (table 4.5). These four 3-d p-algorithms were used to reproduce the first 3 investigations for 3-d assessment space (investigations 4, 5 and 6). It was found that the findings for the 2-d algorithms held for the 3-d algorithms.

The third part of the research question as presented in section 1.5 is: Is there a mathematical system which can be used to construct an independent ordering system for a percentage-based assessment system?

The final investigation, investigation 7, achieved two things. First, an independent ordering system was developed to generate R6 and SAI distributions from percentage-based CBAD. Second, this ordering system was used to produce R6 and SAI distributions for a cohort of senior chemistry students using the four 3-d p-algorithms from investigations 4, 5 and 6. It was found that both R6 and SAI distributions could be generated from the CBAD and that the R6 and SAI distributions for the cohort for each of the 4 OMu were different, reflecting different preference structures. Also, similarities and differences between the R6 and SAIs for the OMu and the school's R6 and SAI distributions were identified. Some of the OMu distributions (e.g. OMu4) were not too different from the school's suggesting the possibility of using either this OMu, or modifying it slightly, to generate R6 and SAIs in the department's assessment system. A more ambitious claim is that the chemistry department may be able to use an independent OS with an appropriate OMu to calculate its R6 and SAI distributions.

This investigation has used the mathematical system developed in section 4.2 to design and apply an independent ordering system for a percentage-based assessment system.

Hence, this study has shown that there is a mathematical system which can be used to construct an independent ordering system for a percentage-based assessment system.

5.5 Conclusions

Two conclusions will be drawn from this study.

First, this study provides a mathematical system for modelling ordering mechanisms for ordering systems used in percentage-based assessment systems. In this way, this study responds to Sadler (1988, p. 10; see section 1.5) who questioned the existence of a mathematical system for this purpose. Although Sadler's question is not actually answered here, it could be pursued by testing the mathematical system developed in this study against the decision structures and ordering mechanisms used in school-based assessment systems. Variations are contained within several of the ordering mechanisms. In OMu1, OMu2 and OMu6 for example, the angle (or angles) of the equivalence curves could vary. For OMu7, the point of origin of the circle could vary. Future exploration of the capacity of ordering mechanisms such as these to model teacher judgements would include the potential 'tuning' of mechanisms to increase the correspondence to teacher's judgements. It is possible that a scaling and fitting exercise undertaken cooperatively by a group of teachers with the guidance of an analyst following the typical processes illustrated throughout by Keeny and Raffia (1976) would address Sadler's basic question about the capacity of some ordering mechanism to effectively model the processes underlying teacher's judgements. Such a study would constitute empirical research into the nature of ordering mechanisms used in school-based Board subject assessment systems.

Added to this, the requirement that SAIs be decisions not calculations (Viviani, 1990, p. 133) is challenged by the findings presented here, since this study suggests that a mathematical system can be developed to model decisions, hence calculations do not necessarily undermine quality decisions, and may actually enhance them. It could be, however, that the prescription against calculation was directed against simplistic a-priori calculations such as simple or weighted averages. This study illustrates the possibility for an approach which requires teacher input to select a modelling system which effectively represents a consensus based decision process. Once (and if) established by application to a representative sample of student results, the application of the model would have the two virtues of consistency and efficiency. Perhaps Viviani's (1990) prescription did not envisage the use of calculations in the context of such modelling.

Secondly, this study has exposed one complex feature of generating ordered distributions of students from multidimensional criteria-based assessment data, namely, that different ordered distributions can be obtained from the same assessment data since the assessment data are not inherently ordered. Different ordering mechanisms yield different ordered distributions. Furthermore, the literature concerned with criteria-based assessment grading does not address this issue since it is primarily concerned with placing profiles of achievement into Levels of Achievement categories, not mapping them onto numerical scales: criteria-based Levels of Achievement are not intervals on a unidimensional continuum. Yet, with the R6 and the SAI distributions, achievement must be expressed on such continuums after the grading process has occurred. Clearly, the issue of generating ordered distributions from such data requires further scrutiny.

BIBLIOGRAPHY

Ackoff, R. L. (1969). Systems, organisations, and interdisciplinary research. In F. E. Emery (Ed.), *Systems thinking* (pp. 330-347). Melbourne: Penguin Australia.

Anastasi, A. (1988). *Psychological testing* (6th ed.). New York: Macmillian.

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review*, *89*, 369-406.

Barnett, V. (1976). The ordering of multivariate data. *Journal of the Royal Statistical Society A*, *139*, 318-355.

Berger, J. O. (1985). *Statistical decision theory and Bayesian analysis* (2nd ed.). New York: Springer-Verlag.

Biggs, J. B. (1991). Good learning: What is it? How can it be fostered?. In J. B. Biggs (Ed.), *Teaching for learning: The view from cognitive psychology* (pp. 215-230). Melbourne: Australian Council for Educational Research.

Biggs, J. B., & Moore, P. J. (1993). *The process of learning* (3rd ed). Sydney: Prentice Hall.

Board of Senior Secondary School Studies. (1988). *The tertiary entrance score – A technical handbook of procedures.* Brisbane: Author.

Board of Senior Secondary School Studies. (1991). *Assessment, moderation and certification in senior secondary school systems in Australia*. Brisbane: Author.

Board of Senior Secondary School Studies. (1992). *Deciding SAIs.* Inservice document prepared for schools by the Board of Senior Secondary School Studies. Brisbane: Author.

Board of Senior Secondary School Studies. (1993a). *Handbook of procedures for accreditation and certification*. Brisbane: Author.

Board of Senior Secondary School Studies. (1993b). *OPs and levels of achievement 1992*. Brisbane: Author.

Bresee, C. W. (1976). On 'grading on the curve'. *Clearing House*, *50*, 108-110.

Campbell, W. J. (1976). *Some consequences of the Radford scheme for schools, teachers and students in Queensland*. Australian Advisory Committee on Research and Development in Education Report No. 7. Canberra: Australian Government Publishing Services.

Chernoff, H. (1973). Using faces to represent points in *k*-dimensional space graphically. *Journal of the American Statistical Association*, *68*, 361-368.

Clarke, E. (1987). *Assessment in Queensland secondary schools: Two decades of change 1964-1983*. Brisbane: Queensland Department of Education.

Cooksey, R. W. (1993). The problem of multidimensionality in course scores and course choices in the production of a single tertiary entrance score. *Australian Journal of Education*, *3*, 26-45.

Cresswell, M. J. (1987). Describing examination performance: Grade criteria in public examinations. *Educational Studies*, *13*, 247-265.

Cronbach, L. J., & Suppes, P. (Eds.). (1969). *Research for tomorrows school's: Disciplined inquiry for education*. New York: Macmillian.

Deutsch, M. (1979). Education and distributive justice: Some reflections on grading systems. *American Psychologist*, *43*, 391-401.

Docking, R. A. (1976). Sum problems in assessment. *Unicorn, 2* (2), 59-64.

Docking, R. A. (1986). Norm-referenced measurement and criterion-referenced measurement: A descriptive comparison. *Unicorn, 12* (1), 40-46.

Docking, R. A. (1987). Constructive credentialling. *Unicorn, 13* (1), 10-17.

Doll, W. E. (1989). Foundations for a post-modern curriculum. *Journal of Curriculum Studies, 21*, 243-253.

Education Act, 1964, Queensland. Government Printing Office.

Education (Senior Secondary School Studies) Act, 1988, Queensland. Government Printing Office.

Fairbairn, K., McBryde, B., & Rigby, D. (1976). *Schools under Radford.* Brisbane: Department of Education, Queensland and Board of Senior Secondary School Studies.

Findlay, J. (1987). *Improving the quality of student performance through assessment*. Discussion paper 15. Brisbane: Board of Senior Secondary School Studies.

Frederiksen, N. (1984). Implications of cognitive theory for instruction in problem solving. *Review of Educational Research*, *54*, 363-407.

Frederiksen, J. R., & Collins, A. (1989). A systems approach to educational testing. *Educational Researcher*, *18* (9), 27-32.

Gough, N. (1989). From epistemology to ecopolitics: Reviewing a paradigm for curriculum. *Journal of Curriculum Studies*, *21*, 225-241.

Hewitson, M. (1988). The concept of performance levels in criterion-referenced assessment. In N. Baumgart (Ed.), *Reports and records of achievement for school leavers* (pp. 240-252). Canberra: The Australian College of Education.

Holmberg, B. (1989). *Theory and practice of distance education*. London: Routledge

Imre, B. W. (1983). Specifications of a grading system. *Higher Education Research and Development*, *2*, 183-196.

Jaeger, R. M. (Ed.). (1988). *Complementary methods for research in education*. Washington, DC: American Educational Research Association.

Judge, H. G. (1985). *Reporting the outcomes of assessment*. Keynote address at the 11th Annual International Association for Educational Assessment Conference, Oxford.

Kamke, E. (1950). *Theory of sets*. New York: Dover.

Keeny, R. L., & Raffia, H. (1976). *Decisions with multiple objectives: Preferences and value tradeoffs*. New York: John Wiley & Sons.

Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. (1971). *Foundations of measurement*. New York: Academic Press.

Larkin, J. H., & Reif, F. (1979). Understanding and teaching problem solving in physics. *European Journal of Science Education*, *1*, 191-203.

Lorand, R. (1992). Bergson's concept of order. *Journal of the History of Philosophy*, *30*, 579-595.

Marton, F. (1981). Phenomenography—Describing conceptions of the world around us. *Instructional Science*, *10*, 177-200.

Marton, F., & Saljo, R. (1976). On qualitative differences in learning—II: Outcome as a function of the learner's conception of the task. *British Journal of Educational Psychology, 46*, 115-127.

Marton, F., & Saljo, R. (1984). Approaches to learning. In F. Marton, D. Hounsell, & N. Entwistle (Eds.), *The experience of learning*. Edinburgh: Scottish Academic Press.

Masters, G. N., & Hill, P. W. (1988). Reforming the assessment of student achievement in the senior secondary school. *Australian Journal of Education, 32*, 274-286.

Meisels, S. J. (1989). High stakes testing. *Educational Leadership, 46* (7), 16-22.

Messick, S. (1984). The psychology of educational measurement. *Journal of Educational Measurement, 21*, 251-237.

Nagel, E. (1969). A formalisation of functionalism. In F. E. Emery (Ed.), *Systems thinking* (pp. 297-329). Melbourne: Penguin Australia.

Naylor, T. H., Burdick, D. S., & Sasser, W. E. (1967). Computer simulation experiments with economic systems: The problem of experimental design. *Journal of the American Statistical Association, 62*, 1315-1337.

Nuthmann, C. (1994). Using human judgement in system dynamics of social systems. *System Dynamics Review, 10*, 1-27.

Ormell, C. (1992). Behavioural objectives revisited. *Educational Research, 34*, 23-33.

Pask, G. (1976). Styles and strategies of learning. *British Journal of Educational Psychology, 46*, 126-148.

Pask, G. (1984). Review of conversation theory and a protologic (or protolanguage), Lp. *ECTJ: Educational Communication and Technology*, *32*, 3-40.

Perry, W. G. (1970). *Forms of intellectual and ethical development in the college years: A scheme*. New York: Holt, Rinehart & Winston.

Power, C. (1986). Criterion-based assessment, grading and reporting at year 12 level. *Australian Journal of Education*, *30*, 266-284.

Radford, W. (Chairman). (1970). *Public examinations for Queensland secondary school students*. Brisbane: Queensland Department of Education.

Richardson, J. T. E. (1987). Research in education and cognitive psychology. In J. T. E. Richardson, M. W. Eysenck, & D. W. Piper (Eds.), *Student learning: Research in education and cognitive psychology* (pp. 3-12). London: The Society for Research into Higher Education & Open University Press.

Rowntree, D. (1977). *Assessing students: How shall we know them*. London: Harper & Row.

Sadler, D. R. (1986a). *ROSBA's family connections*. Discussion paper 1. Brisbane: Board of Senior Secondary School Studies.

Sadler, D. R. (1986b). *General principles for organising criteria*. Discussion paper 9. Brisbane: Board of Senior Secondary School Studies.

Sadler, D. R. (1987). Specifying and promulgating achievement standards. *Oxford Review of Education*, *13*, 191-209.

Sadler, D. R. (1988). *The place of numerical marks in criteria-based assessment*. Discussion paper 21. Brisbane: Board of Senior Secondary School Studies.

Scott, E., Berkeley, G., Howell. M., Schunter, L., Walker, R., & Winkle, L. (1978). *A review of school-based assessment in Queensland secondary schools*. Brisbane: Board of Senior Secondary School Studies.

Scriven, M. (1967). The methodology of evaluation. In R. W. Tyler, R. M. Gagne, & M. Scriven (Eds.), *Perspectives of curriculum evaluation*. Chicago: Rand McNally.

Scriven, M. (1988). Philosophical inquiry methods in education. In R. M. Jaeger (Ed.), *Complementary methods for research in education* (pp. 131-151). Washington, DC: American Educational Research Association.

Secondary Examinations Council (1984). *The development of grade-related criteria for the general certificate of secondary education—A briefing paper for working parties*. London: Author.

Shepard, L. A. (1989). Why we need better assessments. *Educational Leadership*, *46* (7), 4-9.

Shulman, L. S. (1988). Disciplines of inquiry in education: An overview. In R. M. Jaeger (Ed.), *Complementary methods for research in education* (pp. 3-20). Washington, DC: American Educational Research Association.

Slaughter, R. A. (1989). Cultural reconstruction in the post-modern world. *Journal of Curriculum Studies*, *21*, 255-270.

Smith, M. L. (1991). Put to the test: The effects of external testing on teachers. *Educational Researcher*, *20* (5), 8-11.

Smith, D. L., & Lovat, T. J. (1991). *Curriculum action on reflection* (Rev. ed.). Wentworth Falls, NSW: Social Science Press.

Stiggins, R. J., & Bridgeford, N. J. (1985). The ecology of classroom assessment. *Journal of Educational Measurement*, *22*, 271-286.

Taylor, J. C. (1991). Designing instruction to generate expert cognitive skill performance: Empirical evidence. In G. Evans (Ed.), *Learning and teaching cognitive skills* (pp. 164-184). Melbourne: The Australian Council for Educational Research.

Tertiary Entrance Procedures Authority. (1993a). Admission of students into higher education in Australia: A collection of recent research papers. *Research Series* 2. Author.

Tertiary Entrance Procedures Authority. (1993b). *Calculation of OPs and FPs*. Information sheet number 3 August. Author.

Thompson, W. W. (1967). *Operations Research Techniques*. Columbus, OH: Charles E. Merrill.

Van Rossum, E. J., & Schenk, S. M. (1984). The relationship between learning conception, study strategy and learning outcome. *The British Journal of Educational Psychology*, *54*, 73-83.

Viviani, N. (1990). *The review of tertiary entrance in Queensland 1990*. Report submitted to the Minister for Education. Brisbane: Department of Education.

Wilks, S. S. (1963). Multivariate statistical outliers. *Sankhya: The Indian Journal of Statistics Series A*, *25*, 407-426.

Appendix A: Form R6.

Board of Senior Secondary School Studies
Queensland

LEVELS OF ACHIEVEMENT
SCHOOL PROPOSAL

FORM R6
SENIOR
(YEAR 12)

SCHOOL:	School Code: [redacted]
SENIOR SUBJECT	Subject Code: O
DISTRICT:	Panel Code: [redacted]

Relative Achievement of Students

Oct	Panel	Exit	
			Very High Achievement
			High Achievement
			Sound Achievement
			Limited Achievement
			Very Limited Achievement

Entered below are the proposed Levels of Achievement for all eligible students who complete 1, 2, 3 or 4 semesters of the course of study outlined in the school work program.

1. Each Student Folio Sample provided should be made up from the work of a student who is about to exit from the course of study.
2. Each student whose folio is presented, should be identified by a code letter (A, B, C, etc.) on the 50 interval column to indicate relative position within the proposed Level of Achievement.
3. Show the relative achievements of other students, either by placing an 'X' on the column in the appropriate position for each student, or by placing the number of students above or between the sample students in each level.

NO. IN EACH LEVEL OF ACHIEVEMENT

	Very High Achievement	High Achievement	Sound Achievement	Limited Achievement	Very Limited Achievement	TOTAL	SIGNATURE AND DATE
INITIAL SCHOOL PROPOSAL							PRINCIPAL
SUBSEQUENT SCHOOL PROPOSAL							PRINCIPAL
AGREEMENT REACHED ON SCHOOL PROPOSAL							D/S.R.P.Ch
EXIT PROPOSAL							PRINCIPAL
EXIT PROPOSAL RECOMMENDED FOR CERTIFICATION							S.R.P.Ch.

Please tick if the following material has been adequately presented for reviewing.

1. School Work Program. ☐
2. Assessment Instruments. ☐
3. Details of any other method of collecting data on achievement. ☐
4. Conditions sheet for each Assessment Instrument. ☐
5. Samples of Student Work. ☐
6. Relative Achievements of Students. ☐

PANEL COMMENTS: (Please sign and date)

COPIES:

ORIGINAL, DUPLICATE AND TRIPLICATE:	FORWARDED WITH INITIAL AND ANY SUBSEQUENT PROPOSALS.
QUADRUPLICATE	RETAINED BY SCHOOL.
AFTER AGREEMENT ON PROPOSAL:	ORIGINAL AND DUPLICATE RETURNED TO SCHOOL.
TRIPLICATE:	RETAINED BY CHAIRMAN.
EXIT PROPOSAL:	ORIGINAL FORWARDED TO BOARD
	DUPLICATE RETAINED BY SCHOOL.

Appendix B: Program used for investigation 1.

```
REM *****RESEARCH PROGRAM 1*****

CLS

PRINT "RESEARCH PROGRAM 1"


DIM x(100), y(100), p(7, 100)
DIM sx(7, 7), sy(7, 7), sxx(7, 7), syy(7, 7), sxy(7, 7), ssxy(7, 7), ssxx(7, 7), ssyy(7, 7), r(7, 7)


FOR i = 1 TO 100
LET x(i) = 0
LET y(i) = 0
NEXT i


FOR J = 1 TO 7
FOR K = 1 TO 7
LET sx(J, K) = 0: LET sy(J, K) = 0: LET sxx(J, K) = 0: LET syy(J, K) = 0: LET sxy(J, K) = 0
LET ssxy(J, K) = 0: LET ssxx(J, K) = 0: LET ssyy(J, K) = 0
NEXT K
NEXT J


RANDOMIZE
FOR i = 1 TO 100
LET x(i) = INT(RND * 101): LET y(i) = INT(RND * 101)
NEXT i


FOR i = 1 TO 100
LET p(1, i) = (x(i) + y(i)) / 2
LET p(2, i) = (2 * x(i) + y(i)) / (2 + 1)
LET p(3, i) = SQR((x(i) ^ 2 + y(i) ^ 2) / 2)
LET p(4, i) = 100 - SQR((x(i) - 100) ^ 2 + (y(i) - 100) ^ 2) / SQR(2)
IF x(i) >= y(i) THEN LET p(5, i) = y(i) ELSE LET p(5, i) = x(i)
IF x(i) >= y(i) THEN LET p(6, i) = (.5 * x(i) + y(i)) / (.5 + 1) ELSE LET p(6, i) = (2 * x(i) + y(i)) / (2
+ 1)
LET p(7, i) = (200 + 100 - SQR((200 + 100) ^ 2 - 2 * 200 ^ 2 - 2 * 100 ^ 2 + 2 * (x(i) - 200) ^ 2 + 2 *
(y(i) - 100) ^ 2)) / 2
NEXT i


FOR J = 1 TO 7
FOR K = 1 TO 7
FOR i = 1 TO 100
LET sxy(J, K) = sxy(J, K) + p(J, i) * p(K, i)
LET sx(J, K) = sx(J, K) + p(J, i)
```

```
LET sy(J, K) = sy(J, K) + p(K, i)
LET sxx(J, K) = sxx(J, K) + p(J, i) * p(J, i)
LET syy(J, K) = syy(J, K) + p(K, i) * p(K, i)
NEXT i
NEXT K
NEXT J


FOR J = 1 TO 7
FOR K = 1 TO 7
LET ssxy(J, K) = sxy(J, K) - (sx(J, K) * sy(J, K)) / 100
LET ssxx(J, K) = sxx(J, K) - (sx(J, K) * sx(J, K)) / 100
LET ssyy(J, K) = syy(J, K) - (sy(J, K) * sy(J, K)) / 100
LET r(J, K) = INT(1000 * (ssxy(J, K) / SQR(ssxx(J, K) * ssyy(J, K)))) / 1000
NEXT K
NEXT J


FOR J = 1 TO 7
PRINT r(J, 1); r(J, 2); r(J, 3); r(J, 4); r(J, 5); r(J, 6); r(J, 7)
NEXT J

INPUT "ENTER FILE NAME: "; N$
OPEN N$ FOR OUTPUT AS #1

FOR J = 1 TO 7
PRINT #1, r(J, 1), r(J, 2), r(J, 3), r(J, 4), r(J, 5), r(J, 6), r(J, 7)
PRINT #1, ";"

NEXT J

FOR i = 1 TO 100
PRINT #1, i, x(i), y(i), p(1, i), p(2, i), p(3, i), p(4, i), p(5, i), p(6, i), p(7, i)
PRINT #1, ";"
NEXT i
CLOSE
```

Appendix C: The data generated from one trial of RESEARCH PROGRAM 1 for investigation 1.

point	x	y	p1	p2	p3	p4	p5	p6	p7
1	55	59	57	56	57	57	55	56	56
2	67	24	46	53	50	41	24	38	54
3	84	100	92	89	92	89	84	89	85
4	88	63	76	80	77	72	63	71	83
5	85	68	77	79	77	75	68	74	82
6	90	58	74	79	76	69	58	69	83
7	26	93	60	48	68	47	26	48	37
8	98	11	55	69	70	37	11	40	68
9	56	58	57	57	57	57	56	57	56
10	64	51	58	60	58	57	51	55	61
11	48	72	60	56	61	58	48	56	53
12	72	4	38	49	51	29	4	27	49
13	76	52	64	68	65	62	52	60	70
14	73	3	38	50	52	29	3	26	49
15	99	43	71	80	76	60	43	62	85
16	39	9	24	29	28	23	9	19	29
17	93	54	74	80	76	67	54	67	85
18	51	63	57	55	57	57	51	55	54
19	89	22	56	67	65	44	22	44	68
20	2	63	33	22	45	26	2	22	17
21	46	2	24	31	33	21	2	17	31
22	35	69	52	46	55	49	35	46	42
23	44	83	64	57	66	59	44	57	51
24	94	24	59	71	69	46	24	47	73
25	9	19	14	12	15	14	9	12	12
26	12	9	11	11	11	10	9	10	11
27	54	4	29	37	38	25	4	21	37
28	19	20	20	19	20	19	19	19	19
29	50	56	53	52	53	53	50	52	51
30	73	67	70	71	70	70	67	69	72
31	23	27	25	24	25	25	23	24	24
32	13	0	7	9	9	6	0	4	9
33	73	51	62	66	63	60	51	58	68

point	x	y	p1	p2	p3	p4	p5	p6	p7
34	56	68	62	60	62	62	56	60	58
35	18	98	58	45	70	42	18	45	31
36	87	84	86	86	86	85	84	85	87
37	70	72	71	71	71	71	70	71	70
38	4	35	20	14	25	18	4	14	13
39	22	19	21	21	21	20	19	20	21
40	75	8	42	53	53	33	8	30	52
41	75	71	73	74	73	73	71	72	74
42	13	37	25	21	28	24	13	21	20
43	80	4	42	55	57	31	4	29	54
44	94	84	89	91	89	88	84	87	93
45	9	33	21	17	24	20	9	17	16
46	78	37	58	64	61	53	37	51	67
47	52	23	38	42	40	36	23	33	43
48	81	40	61	67	64	55	40	54	70
49	34	27	31	32	31	30	27	29	32
50	45	43	44	44	44	44	43	44	44
51	63	53	58	60	58	58	53	56	61
52	54	78	66	62	67	64	54	62	58
53	99	52	76	83	79	66	52	68	89
54	63	60	62	62	62	61	60	61	62
55	50	30	40	43	41	39	30	37	44
56	57	16	37	43	42	33	16	30	44
57	93	92	93	93	93	92	92	92	93
58	32	36	34	33	34	34	32	33	33
59	15	99	57	43	71	40	15	43	29
60	8	47	28	21	34	25	8	21	18
61	93	62	78	83	79	73	62	72	87
62	55	17	36	42	41	33	17	30	43
63	77	58	68	71	68	66	58	64	73
64	55	81	68	64	69	65	55	64	59
65	17	62	40	32	45	35	17	32	28
66	85	92	89	87	89	88	85	87	86
67	12	17	15	14	15	14	12	14	14

point	x	y	p1	p2	p3	p4	p5	p6	p7
68	20	99	60	46	71	43	20	46	33
69	72	23	48	56	53	42	23	39	57
70	76	46	61	66	63	58	46	56	68
71	18	45	32	27	34	30	18	27	25
72	54	2	28	37	38	23	2	19	36
73	2	60	31	21	42	25	2	21	16
74	35	10	23	27	26	21	10	18	27
75	52	15	34	40	38	31	15	27	40
76	81	64	73	75	73	71	64	70	78
77	42	19	31	34	33	30	19	27	35
78	70	10	40	50	50	33	10	30	50
79	87	59	73	78	74	70	59	68	81
80	54	76	65	61	66	63	54	61	58
81	64	33	49	54	51	46	33	43	55
82	41	16	29	33	31	27	16	24	33
83	58	18	38	45	43	35	18	31	45
84	2	100	51	35	71	31	2	35	19
85	31	62	47	41	49	44	31	41	38
86	8	97	53	38	69	35	8	38	24
87	17	91	54	42	65	41	17	42	30
88	83	18	51	61	60	41	18	40	62
89	45	53	49	48	49	49	45	48	47
90	86	62	74	78	75	71	62	70	81
91	80	82	81	81	81	81	80	81	80
92	26	87	57	46	64	47	26	46	37
93	97	85	91	93	91	89	85	89	96
94	88	19	54	65	64	42	19	42	66
95	74	56	65	68	66	64	56	62	70
96	92	77	85	87	85	83	77	82	90
97	14	18	16	15	16	16	14	15	15
98	79	21	50	60	58	42	21	40	61
99	91	85	88	89	88	88	85	87	90
100	15	33	24	21	26	23	15	21	20

Appendix D: Program used for investigation 2

```
REM *****RESEARCH PROGRAM 2*****

CLS

PRINT "RESEARCH PROGRAM 2"


DIM x(1000), y(1000), p(7, 1000)
DIM sx(7, 7), sy(7, 7), sxx(7, 7), syy(7, 7), sxy(7, 7), ssxy(7, 7), ssxx(7, 7), ssyy(7, 7), r(7, 7)


FOR i = 1 TO 100
LET x(i) = 0
LET y(i) = 0
NEXT i


FOR J = 1 TO 7
FOR k = 1 TO 7
LET sx(J, k) = 0: LET sy(J, k) = 0: LET sxx(J, k) = 0: LET syy(J, k) = 0: LET sxy(J, k) = 0
LET ssxy(J, k) = 0: LET ssxx(J, k) = 0: LET ssyy(J, k) = 0
NEXT k
NEXT J


RANDOMIZE
FOR i = 1 TO 50
LET x(i) = INT(RND * 21): LET y(i) = INT(80 + RND * 21)
NEXT i

FOR i = 51 TO 100
LET x(i) = INT(40 + RND * 21): LET y(i) = INT(40 + RND * 21)
NEXT i


FOR i = 1 TO 100
LET p(1, i) = (x(i) + y(i)) / 2
LET p(2, i) = (2 * x(i) + y(i)) / (2 + 1)
LET p(3, i) = SQR((x(i) ^ 2 + y(i) ^ 2) / 2)
LET p(4, i) = 100 - SQR((x(i) - 100) ^ 2 + (y(i) - 100) ^ 2) / SQR(2)
IF x(i) >= y(i) THEN LET p(5, i) = y(i) ELSE LET p(5, i) = x(i)
IF x(i) >= y(i) THEN LET p(6, i) = (.5 * x(i) + y(i)) / (.5 + 1) ELSE LET p(6, i) = (2 * x(i) + y(i)) / (2
+ 1)
LET p(7, i) = (200 + 100 - SQR((200 + 100) ^ 2 - 2 * 200 ^ 2 - 2 * 100 ^ 2 + 2 * (x(i) - 200) ^ 2 + 2 *
(y(i) - 100) ^ 2)) / 2
NEXT i
```

```
FOR J = 1 TO 7
FOR k = 1 TO 7
FOR i = 1 TO 100
LET sxy(J, k) = sxy(J, k) + p(J, i) * p(k, i)
LET sx(J, k) = sx(J, k) + p(J, i)
LET sy(J, k) = sy(J, k) + p(k, i)
LET sxx(J, k) = sxx(J, k) + p(J, i) * p(J, i)
LET syy(J, k) = syy(J, k) + p(k, i) * p(k, i)
NEXT i
NEXT k
NEXT J


FOR J = 1 TO 7
FOR k = 1 TO 7
LET ssxy(J, k) = sxy(J, k) - (sx(J, k) * sy(J, k)) / 100
LET ssxx(J, k) = sxx(J, k) - (sx(J, k) * sx(J, k)) / 100
LET ssyy(J, k) = syy(J, k) - (sy(J, k) * sy(J, k)) / 100
LET r(J, k) = INT(1000 * (ssxy(J, k) / SQR(ssxx(J, k) * ssyy(J, k)))) / 1000
NEXT k
NEXT J


FOR J = 1 TO 7
PRINT r(J, 1); r(J, 2); r(J, 3); r(J, 4); r(J, 5); r(J, 6); r(J, 7)
NEXT J

INPUT "ENTER FILE NAME: "; N$
OPEN N$ FOR OUTPUT AS #1

FOR J = 1 TO 7
PRINT #1, r(J, 1), r(J, 2), r(J, 3), r(J, 4), r(J, 5), r(J, 6), r(J, 7)
PRINT #1, ";"

NEXT J

FOR i = 1 TO 100
PRINT #1, i, x(i), y(i), p(1, i), p(2, i), p(3, i), p(4, i), p(5, i), p(6, i), p(7, i)
PRINT #1, ";"
NEXT i
CLOSE
```

Appendix E: The data generated from one trial of RESEARCH PROGRAM 2 for investigation 2.

point	x	y	p1	p2	p3	p4	p5	p6	p7
1	19	94	57	44	68	43	19	44	32
2	9	92	51	37	65	35	9	37	24
3	14	89	52	39	64	39	14	39	28
4	6	94	50	35	67	33	6	35	22
5	7	80	44	31	57	33	7	31	22
6	11	97	54	40	69	37	11	40	26
7	20	86	53	42	62	43	20	42	33
8	5	83	44	31	59	32	5	31	21
9	11	97	54	40	69	37	11	40	26
10	19	85	52	41	62	42	19	41	32
11	5	89	47	33	63	32	5	33	21
12	8	89	49	35	63	34	8	35	24
13	4	100	52	36	71	32	4	36	21
14	0	87	44	29	62	29	0	29	17
15	9	90	50	36	64	35	9	36	24
16	10	95	53	38	68	36	10	38	25
17	5	98	52	36	69	33	5	36	21
18	11	83	47	35	59	36	11	35	25
19	5	83	44	31	59	32	5	31	21
20	2	97	50	34	69	31	2	34	19
21	11	91	51	38	65	37	11	38	26
22	9	91	50	36	65	35	9	36	24
23	7	92	50	35	65	34	7	35	23
24	0	98	49	33	69	29	0	33	18
25	8	92	50	36	65	35	8	36	24
26	1	98	50	33	69	30	1	33	18
27	4	82	43	30	58	31	4	30	20
28	18	93	56	43	67	42	18	43	31
29	19	81	50	40	59	41	19	40	31
30	17	90	54	41	65	41	17	41	30
31	6	93	50	35	66	33	6	35	22
32	8	81	45	32	58	34	8	32	23

point	x	y	p1	p2	p3	p4	p5	p6	p7
33	14	97	56	42	69	39	14	42	28
34	13	82	48	36	59	37	13	36	27
35	17	82	50	39	59	40	17	39	30
36	6	99	53	37	70	34	6	37	22
37	3	83	43	30	59	30	3	30	19
38	7	91	49	35	65	34	7	35	23
39	2	81	42	28	57	29	2	28	19
40	8	80	44	32	57	33	8	32	23
41	2	100	51	35	71	31	2	35	19
42	10	99	55	40	70	36	10	40	25
43	2	84	43	29	59	30	2	29	19
44	7	80	44	31	57	33	7	31	22
45	5	93	49	34	66	33	5	34	21
46	10	88	49	36	63	36	10	36	25
47	19	92	56	43	66	42	19	43	32
48	0	94	47	31	66	29	0	31	18
49	15	87	51	39	62	39	15	39	29
50	6	91	49	34	64	33	6	34	22
51	45	53	49	48	49	49	45	48	47
52	42	50	46	45	46	46	42	45	44
53	51	53	52	52	52	52	51	52	51
54	59	58	59	59	59	58	58	58	59
55	41	42	42	41	42	41	41	41	41
56	51	44	48	49	48	47	44	46	49
57	41	56	49	46	49	48	41	46	45
58	57	49	53	54	53	53	49	52	55
59	57	60	59	58	59	58	57	58	58
60	60	43	52	54	52	51	43	49	56
61	52	49	51	51	51	50	49	50	51
62	52	43	48	49	48	47	43	46	50
63	44	55	50	48	50	49	44	48	47
64	49	40	45	46	45	44	40	43	47
65	55	44	50	51	50	49	44	48	52
66	44	49	47	46	47	46	44	46	45

point	x	y	p1	p2	p3	p4	p5	p6	p7
67	51	60	56	54	56	55	51	54	53
68	53	41	47	49	47	47	41	45	50
69	42	56	49	47	49	49	42	47	45
70	56	52	54	55	54	54	52	53	55
71	44	58	51	49	51	51	44	49	47
72	51	58	55	53	55	54	51	53	53
73	54	57	56	55	56	55	54	55	55
74	56	50	53	54	53	53	50	52	55
75	47	55	51	50	51	51	47	50	49
76	60	52	56	57	56	56	52	55	58
77	46	47	47	46	47	46	46	46	46
78	57	57	57	57	57	57	57	57	57
79	43	45	44	44	44	44	43	44	44
80	58	54	56	57	56	56	54	55	57
81	42	43	43	42	43	42	42	42	42
82	41	57	49	46	50	48	41	46	45
83	50	49	50	50	50	49	49	49	50
84	41	44	43	42	43	42	41	42	42
85	47	46	47	47	47	46	46	46	47
86	51	55	53	52	53	53	51	52	52
87	48	55	52	50	52	51	48	50	50
88	53	52	53	53	53	52	52	52	53
89	49	58	54	52	54	53	49	52	51
90	59	40	50	53	50	49	40	46	54
91	50	55	53	52	53	52	50	52	51
92	48	57	53	51	53	52	48	51	50
93	57	41	49	52	50	48	41	46	53
94	58	48	53	55	53	53	48	51	55
95	52	50	51	51	51	51	50	51	52
96	49	42	46	47	46	45	42	44	47
97	46	60	53	51	53	52	46	51	49
98	50	56	53	52	53	53	50	52	51
99	42	46	44	43	44	44	42	43	43
100	52	49	51	51	51	50	49	50	51

Appendix F: The program used for investigation 3.

```
REM *****RESEARCH PROGRAM 3*****

CLS

PRINT "RESEARCH PROGRAM 3"


DIM x(100), y(100), p(7, 100)
DIM sx(7, 7), sy(7, 7), sxx(7, 7), syy(7, 7), sxy(7, 7), ssxy(7, 7), ssxx(7, 7), ssyy(7, 7), r(7, 7)


FOR i = 1 TO 100
LET x(i) = 0
LET y(i) = 0
NEXT i


FOR J = 1 TO 7
FOR k = 1 TO 7
LET sx(J, k) = 0: LET sy(J, k) = 0: LET sxx(J, k) = 0: LET syy(J, k) = 0: LET sxy(J, k) = 0
LET ssxy(J, k) = 0: LET ssxx(J, k) = 0: LET ssyy(J, k) = 0
NEXT k
NEXT J


RANDOMIZE
FOR i = 1 TO 100
LET x(i) = INT(10 + RND * 81): LET y(i) = (x(i) - 10) + INT(RND * 21)
NEXT i


FOR i = 1 TO 100
LET p(1, i) = (x(i) + y(i)) / 2
LET p(2, i) = (2 * x(i) + y(i)) / (2 + 1)
LET p(3, i) = SQR((x(i) ^ 2 + y(i) ^ 2) / 2)
LET p(4, i) = 100 - SQR((x(i) - 100) ^ 2 + (y(i) - 100) ^ 2) / SQR(2)
IF x(i) >= y(i) THEN LET p(5, i) = y(i) ELSE LET p(5, i) = x(i)
IF x(i) >= y(i) THEN LET p(6, i) = (.5 * x(i) + y(i)) / (.5 + 1) ELSE LET p(6, i) = (2 * x(i) + y(i)) / (2
+ 1)
LET p(7, i) = (200 + 100 - SQR((200 + 100) ^ 2 - 2 * 200 ^ 2 - 2 * 100 ^ 2 + 2 * (x(i) - 200) ^ 2 + 2 *
(y(i) - 100) ^ 2)) / 2
NEXT i


FOR J = 1 TO 7
FOR k = 1 TO 7
FOR i = 1 TO 100
LET sxy(J, k) = sxy(J, k) + p(J, i) * p(k, i)
LET sx(J, k) = sx(J, k) + p(J, i)
```

```
LET sy(J, k) = sy(J, k) + p(k, i)
LET sxx(J, k) = sxx(J, k) + p(J, i) * p(J, i)
LET syy(J, k) = syy(J, k) + p(k, i) * p(k, i)
NEXT i
NEXT k
NEXT J


FOR J = 1 TO 7
FOR k = 1 TO 7
LET ssxy(J, k) = sxy(J, k) - (sx(J, k) * sy(J, k)) / 100
LET ssxx(J, k) = sxx(J, k) - (sx(J, k) * sx(J, k)) / 100
LET ssyy(J, k) = syy(J, k) - (sy(J, k) * sy(J, k)) / 100
LET r(J, k) = INT(1000 * (ssxy(J, k) / SQR(ssxx(J, k) * ssyy(J, k)))) / 1000
NEXT k
NEXT J


FOR J = 1 TO 7
PRINT r(J, 1); r(J, 2); r(J, 3); r(J, 4); r(J, 5); r(J, 6); r(J, 7)
NEXT J

INPUT "ENTER FILE NAME: "; N$
OPEN N$ FOR OUTPUT AS #1

FOR J = 1 TO 7
PRINT #1, r(J, 1), r(J, 2), r(J, 3), r(J, 4), r(J, 5), r(J, 6), r(J, 7)
PRINT #1, ";"

NEXT J

FOR i = 1 TO 100
PRINT #1, i, x(i), y(i), p(1, i), p(2, i), p(3, i), p(4, i), p(5, i), p(6, i), p(7, i)
PRINT #1, ";"
NEXT i
CLOSE
```

Appendix G: The data generated from one trial of RESEARCH PROGRAM 3 for investigation 3.

point	x	y	p1	p2	p3	p4	p5	p6	p7
1	52	55	54	53	54	53	52	53	53
2	68	71	70	69	70	69	68	69	69
3	82	72	77	79	77	76	72	75	80
4	27	26	27	27	27	26	26	26	27
5	82	72	77	79	77	76	72	75	80
6	54	58	56	55	56	56	54	55	55
7	69	63	66	67	66	66	63	65	68
8	16	8	12	13	13	12	8	11	13
9	49	54	52	51	52	51	49	51	50
10	38	47	43	41	43	42	38	41	40
11	47	57	52	50	52	52	47	50	49
12	33	26	30	31	30	29	26	28	31
13	84	84	84	84	84	84	84	84	84
14	46	55	51	49	51	50	46	49	48
15	13	23	18	16	19	18	13	16	16
16	61	65	63	62	63	63	61	62	62
17	21	29	25	24	25	25	21	24	23
18	66	69	68	67	68	67	66	67	67
19	17	24	21	19	21	20	17	19	19
20	62	64	63	63	63	63	62	63	62
21	13	17	15	14	15	15	13	14	14
22	87	97	92	90	92	91	87	90	88
23	40	45	43	42	43	42	40	42	41
24	21	31	26	24	26	26	21	24	24
25	82	87	85	84	85	84	82	84	83
26	58	53	56	56	56	55	53	55	57
27	33	23	28	30	28	28	23	26	30
28	64	66	65	65	65	65	64	65	64
29	77	67	72	74	72	72	67	70	75
30	10	18	14	13	15	14	10	13	12
31	39	41	40	40	40	40	39	40	40
32	18	17	18	18	18	17	17	17	18
33	44	41	43	43	43	42	41	42	43

point	x	y	p1	p2	p3	p4	p5	p6	p7
34	19	20	20	19	20	19	19	19	19
35	73	71	72	72	72	72	71	72	73
36	11	5	8	9	9	8	5	7	9
37	17	27	22	20	23	22	17	20	20
38	77	68	73	74	73	72	68	71	75
39	79	89	84	82	84	83	79	82	80
40	33	39	36	35	36	36	33	35	35
41	63	72	68	66	68	67	63	66	65
42	59	60	60	59	60	59	59	59	59
43	55	65	60	58	60	60	55	58	57
44	55	60	58	57	58	57	55	57	56
45	78	78	78	78	78	78	78	78	78
46	81	73	77	78	77	77	73	76	80
47	89	90	90	89	90	89	89	89	89
48	69	78	74	72	74	73	69	72	70
49	72	81	77	75	77	76	72	75	73
50	61	60	61	61	61	60	60	60	61
51	80	70	75	77	75	75	70	73	78
52	48	54	51	50	51	51	48	50	49
53	44	47	46	45	46	45	44	45	45
54	79	89	84	82	84	83	79	82	80
55	17	11	14	15	14	14	11	13	15
56	66	57	62	63	62	61	57	60	64
57	25	16	21	22	21	20	16	19	22
58	13	19	16	15	16	16	13	15	15
59	24	16	20	21	20	20	16	19	21
60	20	13	17	18	17	16	13	15	18
61	39	45	42	41	42	42	39	41	41
62	68	73	71	70	71	70	68	70	69
63	75	68	72	73	72	71	68	70	74
64	67	71	69	68	69	69	67	68	68
65	58	54	56	57	56	56	54	55	57
66	83	79	81	82	81	81	79	80	82
67	30	40	35	33	35	35	30	33	33

point	x	y	p1	p2	p3	p4	p5	p6	p7
68	24	33	29	27	29	28	24	27	27
69	47	46	47	47	47	46	46	46	47
70	64	70	67	66	67	67	64	66	65
71	89	94	92	91	92	91	89	91	89
72	52	50	51	51	51	51	50	51	52
73	83	91	87	86	87	86	83	86	84
74	56	46	51	53	51	51	46	49	53
75	15	19	17	16	17	17	15	16	16
76	51	48	50	50	50	49	48	49	50
77	72	74	73	73	73	73	72	73	72
78	24	29	27	26	27	26	24	26	25
79	69	73	71	70	71	71	69	70	70
80	23	17	20	21	20	20	17	19	21
81	33	29	31	32	31	31	29	30	32
82	58	51	55	56	55	54	51	53	56
83	79	80	80	79	80	79	79	79	79
84	32	23	28	29	28	27	23	26	29
85	58	52	55	56	55	55	52	54	57
86	86	85	86	86	86	85	85	85	86
87	45	52	49	47	49	48	45	47	47
88	83	78	81	81	81	80	78	80	82
89	25	20	23	23	23	22	20	22	23
90	77	72	75	75	75	74	72	74	76
91	19	10	15	16	15	14	10	13	16
92	80	88	84	83	84	84	80	83	81
93	49	53	51	50	51	51	49	50	50
94	84	76	80	81	80	80	76	79	83
95	65	63	64	64	64	64	63	64	65
96	31	22	27	28	27	26	22	25	28
97	32	36	34	33	34	34	32	33	33
98	38	29	34	35	34	33	29	32	35
99	56	46	51	53	51	51	46	49	53
100	85	91	88	87	88	88	85	87	86

Appendix H: Program used for investigation 4.

```
REM *****RESEARCH PROGRAM 4*****

CLS

PRINT "RESEARCH PROGRAM 4"


DIM x(100), y(100), z(100), p(4, 100)
DIM sx(4, 4), sy(4, 4), sxx(4, 4), syy(4, 4), sxy(4, 4), ssxy(4, 4), ssxx(4, 4), ssyy(4, 4), r(4, 4)


REM *****THREE DIMENSIONS*****


PRINT "THREE DIMENSIONS"


FOR J = 1 TO 4
FOR k = 1 TO 4
LET sx(J, k) = 0: LET sy(J, k) = 0: LET sxx(J, k) = 0: LET syy(J, k) = 0: LET sxy(J, k) = 0
LET ssxy(J, k) = 0: LET ssxx(J, k) = 0: LET ssyy(J, k) = 0
NEXT k
NEXT J


FOR i = 1 TO 100
LET x(i) = 0
LET y(i) = 0
LET z(i) = 0
NEXT i

RANDOMIZE


FOR i = 1 TO 100
LET x(i) = INT(RND * 101): LET y(i) = INT(RND * 101): LET z(i) = INT(RND * 101)
NEXT i


FOR i = 1 TO 100
LET p(1, i) = (x(i) + y(i) + z(i)) / 3
LET p(2, i) = SQR((x(i) ^ 2 + y(i) ^ 2 + z(i) ^ 2) / 3)
LET p(3, i) = 100 - SQR((x(i) - 100) ^ 2 + (y(i) - 100) ^ 2 + (z(i) - 100) ^ 2) / SQR(3)
IF x(i) <= y(i) AND x(i) <= z(i) THEN LET p(4, i) = x(i)
IF y(i) <= x(i) AND y(i) <= z(i) THEN LET p(4, i) = y(i)
IF z(i) <= x(i) AND z(i) <= y(i) THEN LET p(4, i) = z(i)
NEXT i
```

```
FOR J = 1 TO 4
FOR k = 1 TO 4
FOR i = 1 TO 100
LET sxy(J, k) = sxy(J, k) + p(J, i) * p(k, i)
LET sx(J, k) = sx(J, k) + p(J, i)
LET sy(J, k) = sy(J, k) + p(k, i)
LET sxx(J, k) = sxx(J, k) + p(J, i) * p(J, i)
LET syy(J, k) = syy(J, k) + p(k, i) * p(k, i)
NEXT i
NEXT k
NEXT J


FOR J = 1 TO 4
FOR k = 1 TO 4
LET ssxy(J, k) = sxy(J, k) - (sx(J, k) * sy(J, k)) / 100
LET ssxx(J, k) = sxx(J, k) - (sx(J, k) * sx(J, k)) / 100
LET ssyy(J, k) = syy(J, k) - (sy(J, k) * sy(J, k)) / 100
LET r(J, k) = INT(1000 * (ssxy(J, k) / SQR(ssxx(J, k) * ssyy(J, k)))) / 1000
NEXT k
NEXT J


FOR J = 1 TO 4
PRINT r(J, 1), r(J, 2), r(J, 3), r(J, 4)
NEXT J


INPUT "ENTER FILE NAME: "; N$
OPEN N$ FOR OUTPUT AS #1

FOR J = 1 TO 4
PRINT #1, r(J, 1), r(J, 2), r(J, 3), r(J, 4)
PRINT #1, ";"
NEXT J

FOR i = 1 TO 100
PRINT #1, i, x(i), y(i), z(i), p(1, i), p(2, i), p(3, i), p(4, i)
PRINT #1, ";"
NEXT i
CLOSE
```

Appendix I: The data generated from one trial of RESEARCH PROGRAM 4 for investigation 4.

point	x	y	z	p1	p3	p4	p5
1	56	78	42	59	61	56	42
2	8	99	53	53	65	40	8
3	31	16	66	38	43	34	16
4	100	77	58	78	80	72	58
5	44	87	71	67	70	63	44
6	27	1	44	24	30	22	1
7	35	46	31	37	38	37	31
8	22	23	36	27	28	27	22
9	22	85	53	53	59	47	22
10	23	57	18	33	37	30	18
11	67	62	26	52	55	48	26
12	77	13	87	59	67	48	13
13	83	40	55	59	62	56	40
14	92	2	7	34	53	22	2
15	2	28	85	38	52	29	2
16	9	70	33	37	45	32	9
17	75	47	59	60	61	59	47
18	83	3	57	48	58	38	3
19	67	65	60	64	64	64	60
20	0	20	84	35	50	25	0
21	91	74	31	65	70	57	31
22	85	13	53	50	58	42	13
23	79	7	66	51	60	42	7
24	54	22	62	46	49	43	22
25	77	24	27	43	49	38	24
26	28	60	53	47	49	45	28
27	28	86	34	49	56	43	28
28	15	8	63	29	38	25	8
29	70	33	11	38	45	33	11
30	15	57	63	45	50	41	15
31	69	25	4	33	42	27	4
32	14	63	30	36	41	33	14
33	56	84	4	48	58	38	4

point	x	y	z	p1	p3	p4	p5
34	74	39	23	45	50	41	23
35	48	81	31	53	57	49	31
36	28	15	62	35	40	32	15
37	60	47	62	56	57	56	47
38	38	21	28	29	30	29	21
39	51	91	21	54	61	46	21
40	79	73	41	64	66	61	41
41	54	52	64	57	57	56	52
42	52	89	69	70	72	66	52
43	77	52	97	75	78	69	52
44	48	56	88	64	66	60	48
45	93	77	48	73	75	67	48
46	1	5	97	34	56	21	1
47	34	32	76	47	51	44	32
48	19	86	42	49	56	42	19
49	76	64	53	64	65	63	53
50	67	49	21	46	49	42	21
51	66	95	91	84	85	79	66
52	46	71	67	61	62	60	46
53	34	66	91	64	68	57	34
54	2	75	23	33	45	27	2
55	99	53	16	56	65	44	16
56	41	38	77	52	55	49	38
57	40	9	47	32	36	30	9
58	42	99	98	80	84	66	42
59	16	52	38	35	38	34	16
60	100	78	97	92	92	87	78
61	43	91	6	47	58	36	6
62	30	8	21	20	22	19	8
63	57	73	31	54	56	51	31
64	32	17	35	28	29	28	17
65	32	89	99	73	79	60	32
66	73	3	46	41	50	34	3
67	34	61	76	57	60	54	34

point	x	y	z	p1	p3	p4	p5
68	43	21	95	53	61	44	21
69	17	10	70	32	42	27	10
70	20	49	13	27	31	26	13
71	94	82	95	90	91	89	82
72	12	11	100	41	58	28	11
73	48	58	86	64	66	61	48
74	73	70	98	80	81	77	70
75	75	3	7	28	44	21	3
76	20	20	50	30	33	29	20
77	42	18	52	37	40	36	18
78	5	54	76	45	54	38	5
79	21	94	24	46	57	37	21
80	86	60	1	49	61	38	1
81	50	84	64	66	67	63	50
82	27	94	17	46	57	36	17
83	96	57	49	67	70	61	49
84	33	23	70	42	47	39	23
85	41	90	78	70	73	63	41
86	59	43	99	67	71	59	43
87	30	88	23	47	55	40	23
88	87	6	64	52	62	41	6
89	14	88	13	38	52	29	13
90	41	8	51	33	38	31	8
91	4	47	81	44	54	36	4
92	89	55	34	59	64	53	34
93	30	31	74	45	49	41	30
94	56	53	24	44	47	42	24
95	37	68	37	47	50	45	37
96	36	42	57	45	46	44	36
97	80	0	95	58	72	41	0
98	90	39	80	70	73	62	39
99	12	90	98	67	77	49	12
100	54	64	40	53	54	52	40

Appendix J: Program used for investigation 5.

```
REM *****RESEARCH PROGRAM 5*****

CLS

PRINT "RESEARCH PROGRAM 5"


DIM x(100), y(100), z(100), p(4, 100)
DIM sx(4, 4), sy(4, 4), sxx(4, 4), syy(4, 4), sxy(4, 4), ssxy(4, 4), ssxx(4, 4), ssyy(4, 4), r(4, 4)


REM *****THREE DIMENSIONS*****


PRINT "THREE DIMENSIONS"


FOR J = 1 TO 4
FOR k = 1 TO 4
LET sx(J, k) = 0: LET sy(J, k) = 0: LET sxx(J, k) = 0: LET syy(J, k) = 0: LET sxy(J, k) = 0
LET ssxy(J, k) = 0: LET ssxx(J, k) = 0: LET ssyy(J, k) = 0
NEXT k
NEXT J


FOR i = 1 TO 100
LET x(i) = 0
LET y(i) = 0
LET z(i) = 0
NEXT i

RANDOMIZE


FOR i = 1 TO 50
LET x(i) = INT(40 + RND * 21): LET y(i) = INT(40 + RND * 21): LET z(i) = INT(40 + RND * 21)
NEXT i

FOR i = 51 TO 100
LET x(i) = INT(RND * 21): LET y(i) = INT(RND * 21): LET z(i) = INT(80 + RND * 21)
NEXT i

FOR i = 1 TO 100
LET p(1, i) = (x(i) + y(i) + z(i)) / 3
LET p(2, i) = SQR((x(i) ^ 2 + y(i) ^ 2 + z(i) ^ 2) / 3)
LET p(3, i) = 100 - SQR((x(i) - 100) ^ 2 + (y(i) - 100) ^ 2 + (z(i) - 100) ^ 2) / SQR(3)
IF x(i) <= y(i) AND x(i) <= z(i) THEN LET p(4, i) = x(i)
IF y(i) <= x(i) AND y(i) <= z(i) THEN LET p(4, i) = y(i)
IF z(i) <= x(i) AND z(i) <= y(i) THEN LET p(4, i) = z(i)
```

```
NEXT i

FOR J = 1 TO 4
FOR k = 1 TO 4
FOR i = 1 TO 100
LET sxy(J, k) = sxy(J, k) + p(J, i) * p(k, i)
LET sx(J, k) = sx(J, k) + p(J, i)
LET sy(J, k) = sy(J, k) + p(k, i)
LET sxx(J, k) = sxx(J, k) + p(J, i) * p(J, i)
LET syy(J, k) = syy(J, k) + p(k, i) * p(k, i)
NEXT i
NEXT k
NEXT J

FOR J = 1 TO 4
FOR k = 1 TO 4
LET ssxy(J, k) = sxy(J, k) - (sx(J, k) * sy(J, k)) / 100
LET ssxx(J, k) = sxx(J, k) - (sx(J, k) * sx(J, k)) / 100
LET ssyy(J, k) = syy(J, k) - (sy(J, k) * sy(J, k)) / 100
LET r(J, k) = INT(1000 * (ssxy(J, k) / SQR(ssxx(J, k) * ssyy(J, k)))) / 1000
NEXT k
NEXT J

FOR J = 1 TO 4
PRINT r(J, 1), r(J, 2), r(J, 3), r(J, 4)
NEXT J

INPUT "ENTER FILE NAME: "; N$
OPEN N$ FOR OUTPUT AS #1

FOR J = 1 TO 4
PRINT #1, r(J, 1), r(J, 2), r(J, 3), r(J, 4)
PRINT #1, ";"
NEXT J

FOR i = 1 TO 100
PRINT #1, i, x(i), y(i), z(i), p(1, i), p(2, i), p(3, i), p(4, i)
PRINT #1, ";"
NEXT i
CLOSE
```

Appendix K: The data generated from one trial of RESEARCH PROGRAM 5 for investigation 5.

point	x	y	z	p1	p3	p4	p5
1	46	53	44	48	48	48	44
2	48	59	58	55	55	55	48
3	53	58	49	53	53	53	49
4	57	54	51	54	54	54	51
5	50	59	60	56	57	56	50
6	41	50	43	45	45	45	41
7	52	50	43	48	48	48	43
8	58	54	55	56	56	56	54
9	41	48	45	45	45	45	41
10	56	58	48	54	54	54	48
11	49	59	44	51	51	50	44
12	42	51	47	47	47	47	42
13	46	42	58	49	49	48	42
14	53	55	42	50	50	50	42
15	46	42	44	44	44	44	42
16	44	55	59	53	53	52	44
17	59	43	59	54	54	53	43
18	50	59	41	50	51	49	41
19	52	51	58	54	54	54	51
20	45	56	56	52	53	52	45
21	40	51	47	46	46	46	40
22	49	45	46	47	47	47	45
23	57	54	54	55	55	55	54
24	52	51	48	50	50	50	48
25	48	50	59	52	53	52	48
26	40	45	55	47	47	46	40
27	41	44	43	43	43	43	41
28	57	51	60	56	56	56	51
29	53	54	51	53	53	53	51
30	51	46	53	50	50	50	46
31	42	52	60	51	52	51	42
32	51	58	43	51	51	50	43
33	58	42	54	51	52	51	42

point	x	y	z	p1	p3	p4	p5
34	50	60	52	54	54	54	50
35	56	52	58	55	55	55	52
36	55	43	45	48	48	47	43
37	51	56	42	50	50	49	42
38	50	55	59	55	55	55	50
39	41	58	44	48	48	47	41
40	60	53	47	53	54	53	47
41	54	44	60	53	53	52	44
42	45	53	53	50	50	50	45
43	45	54	49	49	49	49	45
44	44	57	56	52	53	52	44
45	41	50	56	49	49	49	41
46	57	51	53	54	54	54	51
47	55	58	40	51	52	50	40
48	52	42	46	47	47	47	42
49	60	44	44	49	50	49	44
50	46	51	44	47	47	47	44
51	1	8	90	33	52	22	1
52	18	18	96	44	57	33	18
53	18	4	93	38	55	27	4
54	9	8	88	35	51	25	8
55	18	12	98	43	58	31	12
56	9	14	92	38	54	28	9
57	17	20	84	40	51	33	17
58	18	11	89	39	53	30	11
59	3	4	85	31	49	21	3
60	11	12	85	36	50	27	11
61	11	15	100	42	59	29	11
62	5	2	87	31	50	21	2
63	12	5	93	37	54	25	5
64	9	3	81	31	47	22	3
65	3	11	89	34	52	24	3
66	19	6	97	41	57	28	6
67	13	9	82	35	48	27	9

point	x	y	z	p1	p3	p4	p5
68	11	6	86	34	50	25	6
69	1	14	89	35	52	24	1
70	13	1	93	36	54	24	1
71	17	15	82	38	49	31	15
72	16	10	90	39	53	29	10
73	19	8	96	41	57	29	8
74	13	17	94	41	56	30	13
75	20	0	82	34	49	25	0
76	6	6	98	37	57	23	6
77	5	19	96	40	57	28	5
78	19	17	81	39	49	32	17
79	14	3	88	35	51	25	3
80	10	1	85	32	49	22	1
81	15	6	94	38	55	27	6
82	17	3	94	38	55	26	3
83	17	12	87	39	52	30	12
84	18	0	80	33	47	24	0
85	7	4	85	32	49	22	4
86	17	13	91	40	54	30	13
87	0	8	92	33	53	21	0
88	19	6	91	39	54	28	6
89	7	8	100	38	58	24	7
90	11	11	85	36	50	27	11
91	5	9	95	36	55	24	5
92	6	14	91	37	53	26	6
93	12	14	82	36	49	28	12
94	2	17	90	36	53	26	2
95	0	8	82	30	48	21	0
96	5	2	81	29	47	20	2
97	3	0	85	29	49	19	0
98	11	6	83	33	48	25	6
99	8	17	82	36	49	28	8
100	5	17	89	37	52	27	5

Appendix L: Program used for investigation 6.

```
REM *****RESEARCH PROGRAM 6*****

CLS

PRINT "RESEARCH PROGRAM 6"


DIM x(100), y(100), z(100), p(4, 100)
DIM sx(4, 4), sy(4, 4), sxx(4, 4), syy(4, 4), sxy(4, 4), ssxy(4, 4), ssxx(4, 4), ssyy(4, 4), r(4, 4)


REM *****THREE DIMENSIONS*****


PRINT "THREE DIMENSIONS"


FOR J = 1 TO 4
FOR k = 1 TO 4
LET sx(J, k) = 0: LET sy(J, k) = 0: LET sxx(J, k) = 0: LET syy(J, k) = 0: LET sxy(J, k) = 0
LET ssxy(J, k) = 0: LET ssxx(J, k) = 0: LET ssyy(J, k) = 0
NEXT k
NEXT J


FOR i = 1 TO 100
LET x(i) = 0
LET y(i) = 0
LET z(i) = 0
NEXT i

RANDOMIZE


FOR i = 1 TO 100
LET x(i) = INT(10 + RND * 81): LET y(i) = (x(i) - 10) + INT(RND * 21): LET z(i) = (x(i) - 10) +
INT(RND * 21)
NEXT i


FOR i = 1 TO 100
LET p(1, i) = (x(i) + y(i) + z(i)) / 3
LET p(2, i) = SQR((x(i) ^ 2 + y(i) ^ 2 + z(i) ^ 2) / 3)
LET p(3, i) = 100 - SQR((x(i) - 100) ^ 2 + (y(i) - 100) ^ 2 + (z(i) - 100) ^ 2) / SQR(3)
IF x(i) <= y(i) AND x(i) <= z(i) THEN LET p(4, i) = x(i)
IF y(i) <= x(i) AND y(i) <= z(i) THEN LET p(4, i) = y(i)
IF z(i) <= x(i) AND z(i) <= y(i) THEN LET p(4, i) = z(i)
NEXT i
```

```
FOR J = 1 TO 4
FOR k = 1 TO 4
FOR i = 1 TO 100
LET sxy(J, k) = sxy(J, k) + p(J, i) * p(k, i)
LET sx(J, k) = sx(J, k) + p(J, i)
LET sy(J, k) = sy(J, k) + p(k, i)
LET sxx(J, k) = sxx(J, k) + p(J, i) * p(J, i)
LET syy(J, k) = syy(J, k) + p(k, i) * p(k, i)
NEXT i
NEXT k
NEXT J


FOR J = 1 TO 4
FOR k = 1 TO 4
LET ssxy(J, k) = sxy(J, k) - (sx(J, k) * sy(J, k)) / 100
LET ssxx(J, k) = sxx(J, k) - (sx(J, k) * sx(J, k)) / 100
LET ssyy(J, k) = syy(J, k) - (sy(J, k) * sy(J, k)) / 100
LET r(J, k) = INT(1000 * (ssxy(J, k) / SQR(ssxx(J, k) * ssyy(J, k)))) / 1000
NEXT k
NEXT J


FOR J = 1 TO 4
PRINT r(J, 1), r(J, 2), r(J, 3), r(J, 4)
NEXT J


INPUT "ENTER FILE NAME: "; N$
OPEN N$ FOR OUTPUT AS #1

FOR J = 1 TO 4
PRINT #1, r(J, 1), r(J, 2), r(J, 3), r(J, 4)
PRINT #1, ";"
NEXT J

FOR i = 1 TO 100
PRINT #1, i, x(i), y(i), z(i), p(1, i), p(2, i), p(3, i), p(4, i)
PRINT #1, ";"
NEXT i
CLOSE
```

Appendix M: The data generated from one trial of RESEARCH PROGRAM 6 for investigation 6.

point	x	y	z	p1	p3	p4	p5
1	63	66	59	63	63	63	59
2	70	75	78	74	74	74	70
3	35	26	29	30	30	30	26
4	50	55	60	55	55	55	50
5	90	98	95	94	94	93	90
6	14	23	22	20	20	20	14
7	58	50	49	52	52	52	49
8	63	73	54	63	64	63	54
9	42	42	35	40	40	40	35
10	68	76	75	73	73	73	68
11	71	66	73	70	70	70	66
12	81	77	85	81	81	81	77
13	23	13	26	21	21	20	13
14	53	52	62	56	56	55	52
15	85	95	87	89	89	88	85
16	51	50	61	54	54	54	50
17	22	20	19	20	20	20	19
18	82	84	83	83	83	83	82
19	70	66	79	72	72	71	66
20	29	19	36	28	29	28	19
21	65	56	62	61	61	61	56
22	56	50	53	53	53	53	50
23	72	73	63	69	69	69	63
24	89	98	88	92	92	91	88
25	56	66	58	60	60	60	56
26	69	59	77	68	69	67	59
27	31	38	36	35	35	35	31
28	15	23	17	18	19	18	15
29	56	53	46	52	52	51	46
30	31	31	41	34	35	34	31
31	22	17	28	22	23	22	17
32	61	54	63	59	59	59	54
33	84	93	80	86	86	85	80

34	37	34	47	39	40	39	34
35	32	29	42	34	35	34	29
36	40	38	37	38	38	38	37
37	28	23	24	25	25	25	23
38	23	15	27	22	22	22	15
39	79	74	88	80	81	79	74
40	39	29	39	36	36	35	29
41	58	58	53	56	56	56	53
42	14	5	22	14	15	13	5
43	32	41	23	32	33	32	23
44	34	36	26	32	32	32	26
45	87	83	95	88	88	87	83
46	71	63	80	71	72	71	63
47	89	96	80	88	89	87	80
48	82	88	83	84	84	84	82
49	64	54	73	64	64	63	54
50	85	85	92	87	87	87	85
51	73	73	75	74	74	74	73
52	53	63	47	54	55	54	47
53	45	35	53	44	45	44	35
54	90	82	86	86	86	86	82
55	36	45	42	41	41	41	36
56	31	39	25	32	32	31	25
57	37	35	27	33	33	33	27
58	24	24	16	21	22	21	16
59	76	75	82	78	78	77	75
60	79	83	81	81	81	81	79
61	30	21	31	27	28	27	21
62	73	82	82	79	79	79	73
63	10	16	17	14	15	14	10
64	38	30	36	35	35	35	30
65	45	49	37	44	44	43	37
66	71	75	63	70	70	69	63
67	44	39	35	39	40	39	35
68	60	53	60	58	58	58	53

69	40	33	41	38	38	38	33
70	49	58	59	55	56	55	49
71	81	78	72	77	77	77	72
72	39	38	48	42	42	41	38
73	85	83	95	88	88	87	83
74	76	69	69	71	71	71	69
75	23	27	22	24	24	24	22
76	29	31	33	31	31	31	29
77	62	54	66	61	61	60	54
78	43	48	39	43	43	43	39
79	16	21	19	19	19	19	16
80	36	30	28	31	32	31	28
81	45	38	37	40	40	40	37
82	27	35	37	33	33	33	27
83	60	59	51	57	57	56	51
84	70	70	66	69	69	69	66
85	37	43	30	37	37	36	30
86	35	36	41	37	37	37	35
87	50	54	51	52	52	52	50
88	60	67	54	60	61	60	54
89	68	64	66	66	66	66	64
90	88	81	97	89	89	87	81
91	79	88	84	84	84	83	79
92	27	18	20	22	22	22	18
93	56	49	51	52	52	52	49
94	83	88	89	87	87	86	83
95	33	29	27	30	30	30	27
96	43	39	51	44	45	44	39
97	24	14	22	20	20	20	14
98	80	75	87	81	81	80	75
99	10	3	9	7	8	7	3
100	35	42	29	35	36	35	29

Appendix N: Program used for investigation 7.

```
REM *****RESEARCH PROGRAM 7*****

CLS

PRINT "RESEARCH PROGRAM 7"


DIM x(100), y(100), z(100), a(100), op(7, 100), sai(5, 100)
DIM sx(5, 5), sy(5, 5), sxx(5, 5), syy(5, 5), sxy(5, 5), ssxy(5, 5), ssxx(5, 5), ssyy(5, 5), r(5, 5)
DIM oploa(7, 10), loa(7, 10), oploadiv(7, 10), oprungdiv(7, 100), oprung(7, 100)

FOR j = 1 TO 5
FOR k = 1 TO 5
LET sx(j, k) = 0: LET sy(j, k) = 0: LET sxx(j, k) = 0: LET syy(j, k) = 0: LET sxy(j, k) = 0
LET ssxy(j, k) = 0: LET ssxx(j, k) = 0: LET ssyy(j, k) = 0
NEXT k
NEXT j

PRINT "INPUT THE NUMBER OF STUDENTS IN THE COHORT"
INPUT n


FOR i = 1 TO n
LET x(i) = 0
LET y(i) = 0
LET z(i) = 0
NEXT i


FOR i = 1 TO n
PRINT "STUDENT "; i
INPUT "X VALUE"; x(i)
INPUT "Y VALUE"; y(i)
INPUT "Z VALUE"; z(i)
NEXT i

FOR i = 1 TO n
PRINT "student "; i
INPUT "SAI "; sai(5, i)
NEXT i

FOR i = 1 TO n
LET op(1, i) = (x(i) + y(i) + z(i)) / 3
LET op(2, i) = SQR((x(i) ^ 2 + y(i) ^ 2 + z(i) ^ 2) / 3)
```

```
LET op(3, i) = 100 - SQR((x(i) - 100) ^ 2 + (y(i) - 100) ^ 2 + (z(i) - 100) ^ 2) / SQR(3)
IF x(i) <= y(i) AND x(i) <= z(i) THEN LET op(4, i) = x(i)
IF y(i) <= x(i) AND y(i) <= z(i) THEN LET op(4, i) = y(i)
IF z(i) <= x(i) AND z(i) <= y(i) THEN LET op(4, i) = z(i)
NEXT i

FOR j = 1 TO 4
LET min(j) = 100
LET max(j) = 0
NEXT j


FOR j = 1 TO 4
FOR k = 1 TO n
IF op(j, k) <= min(j) THEN LET min(j) = op(j, k)
IF op(j, k) >= max(j) THEN LET max(j) = op(j, k)
NEXT k
NEXT j

FOR j = 1 TO 4
FOR k = 1 TO n
LET sai(j, k) = INT(((op(j, k) - min(j)) / (max(j) - min(j)) * 200) * 100) / 100 + 200
NEXT k
NEXT j


FOR j = 1 TO 4
FOR k = 1 TO 4
FOR i = 1 TO n
LET sxy(j, k) = sxy(j, k) + op(j, i) * op(k, i)
LET sx(j, k) = sx(j, k) + op(j, i)
LET sy(j, k) = sy(j, k) + op(k, i)
LET sxx(j, k) = sxx(j, k) + op(j, i) * op(j, i)
LET syy(j, k) = syy(j, k) + op(k, i) * op(k, i)
NEXT i
NEXT k
NEXT j


FOR j = 1 TO 4
FOR k = 1 TO 4
LET ssxy(j, k) = sxy(j, k) - (sx(j, k) * sy(j, k)) / n
LET ssxx(j, k) = sxx(j, k) - (sx(j, k) * sx(j, k)) / n
LET ssyy(j, k) = syy(j, k) - (sy(j, k) * sy(j, k)) / n
LET r(j, k) = INT(1000 * (ssxy(j, k) / SQR(ssxx(j, k) * ssyy(j, k)))) / 1000
NEXT k
```

```
NEXT j


FOR j = 1 TO 4
PRINT r(j, 1), r(j, 2), r(j, 3), r(j, 4)
NEXT j


INPUT "ENTER FILE NAME: "; n$
OPEN n$ FOR OUTPUT AS #1

FOR j = 1 TO 4
PRINT #1, r(j, 1), r(j, 2), r(j, 3), r(j, 4)
PRINT #1, ";"
NEXT j


REM SAI CORRELATION

FOR j = 1 TO 5
FOR k = 1 TO 5
LET sx(j, k) = 0: LET sy(j, k) = 0: LET sxx(j, k) = 0: LET syy(j, k) = 0: LET sxy(j, k) = 0
LET ssxy(j, k) = 0: LET ssxx(j, k) = 0: LET ssyy(j, k) = 0
NEXT k
NEXT j


FOR j = 1 TO 5
FOR k = 1 TO 5
FOR i = 1 TO n
LET sxy(j, k) = sxy(j, k) + sai(j, i) * sai(k, i)
LET sx(j, k) = sx(j, k) + sai(j, i)
LET sy(j, k) = sy(j, k) + sai(k, i)
LET sxx(j, k) = sxx(j, k) + sai(j, i) * sai(j, i)
LET syy(j, k) = syy(j, k) + sai(k, i) * sai(k, i)
NEXT i
NEXT k
NEXT j


FOR j = 1 TO 5
FOR k = 1 TO 5
LET ssxy(j, k) = sxy(j, k) - (sx(j, k) * sy(j, k)) / n
LET ssxx(j, k) = sxx(j, k) - (sx(j, k) * sx(j, k)) / n
LET ssyy(j, k) = syy(j, k) - (sy(j, k) * sy(j, k)) / n
LET r(j, k) = INT(1000 * (ssxy(j, k) / SQR(ssxx(j, k) * ssyy(j, k)))) / 1000
```

```
NEXT k
NEXT j


FOR j = 1 TO 5
PRINT r(j, 1), r(j, 2), r(j, 3), r(j, 4), r(j, 5)
NEXT j

PRINT #1, ";"

FOR j = 1 TO 5
PRINT #1, r(j, 1), r(j, 2), r(j, 3), r(j, 4), r(j, 5)
PRINT #1, ";"
NEXT j


REM CALCULATION OF FORM R6

PRINT "ENTER CUT-OFF MARKS FOR EACH LEVEL OF ACHIEVEMENT"

INPUT "VHA X "; loa(1, 1)
INPUT "VHA Y "; loa(1, 2)
INPUT "VHA Z "; loa(1, 3)
PRINT
INPUT "HA X "; loa(2, 1)
INPUT "HA Y "; loa(2, 2)
INPUT "HA Z "; loa(2, 3)
PRINT
INPUT "SA X "; loa(3, 1)
INPUT "SA Y "; loa(3, 2)
INPUT "SA Z "; loa(3, 3)
PRINT
INPUT "LA X "; loa(4, 1)
INPUT "LA Y "; loa(4, 2)
INPUT "LA Z "; loa(4, 3)
PRINT
INPUT "VLA X "; loa(5, 1)
INPUT "VLA Y "; loa(5, 2)
INPUT "VLA Z "; loa(5, 3)


FOR i = 1 TO 5
LET oploa(1, i) = (loa(i, 1) + loa(i, 2) + loa(i, 3)) / 3
LET oploa(2, i) = SQR((loa(i, 1) ^ 2 + loa(i, 2) ^ 2 + loa(i, 3) ^ 2) / 3)
LET oploa(3, i) = 100 - SQR((loa(i, 1) - 100) ^ 2 + (loa(i, 2) - 100) ^ 2 + (loa(i, 3) - 100) ^ 2)
/ SQR(3)
```

```
IF loa(i,1) <= loa(i,2) AND loa(i,1) <= loa(i,3) THEN LET oploa(4,i) = loa(i,1)
IF loa(i,2) <= loa(i,1) AND loa(i,2) <= loa(i,3) THEN LET oploa(4, i) = loa(i,2)
IF loa(i,3) <= loa(i,1) AND loa(i,3) <= loa(i,2) THEN LET oploa(4,i) = loa(i,3)
NEXT i

FOR k = 1 TO 4
LET oploadiv(k, 1) = (100 - oploa(k, 1)) / 10
NEXT k

FOR j = 1 TO 4
FOR i = 2 TO 5
LET oploadiv(j, i) = (oploa(j, i - 1) - oploa(j, i)) / 10
NEXT i
NEXT j



FOR k = 1 TO 4
FOR j = 0 TO 4
FOR i = 1 TO 10
LET oprungdiv(k, j * 10 + i) = oploa(k, (5 - j)) + i * oploadiv(k, (5 - j))
NEXT i
NEXT j
NEXT k

FOR k = 1 TO 4
FOR j = 1 TO n
FOR i = 1 TO 50

IF op(k, j) < oprungdiv(k, i) THEN LET oprung(k, j) = i: GOTO 10


NEXT i
10 NEXT j
NEXT k

PRINT "HERE IS THE R6 INFORMATION"

FOR i = 1 TO n
PRINT x(i), y(i), z(i)
FOR j = 1 TO 4
PRINT oprung(j, i); sai(j, i)
NEXT j
NEXT i
```

```
FOR i = 1 TO n
PRINT #1, i, x(i), y(i), z(i), sai(5, i), op(1, i), op(2, i), op(3, i), op(4, i), sai(1, i), sai(2, i),
sai(3, i), sai(4, i), oprung(1, i), oprung(2, i), oprung(3, i), oprung(4, i)
PRINT #1, ";"
NEXT i
CLOSE
```

Appendix O:The data generated from one trial of RESEARCH PROGRAM 7 for investigation 7.

School data																
S.No.	x	y	z	sai	r6	p1	p3	p4	p5	saip1	saip3	saip4	saip5	r6p1	r6p3	r6p4	r6p5
1	60	50	25	246	20	45	47	43	25	253	255	250	229	21	22	19	9
2	59	28	45	234	17	44	46	43	28	250	250	248	237	20	21	19	10
3	84	79	75	351	42	79	79	79	75	351	349	351	351	40	40	40	37
4	87	69	68	317	35	75	75	73	68	337	337	335	334	37	37	36	33
5	55	58	55	276	26	56	56	56	55	284	280	286	302	26	26	26	26
6	75	52	55	271	25	61	62	59	52	298	296	296	295	28	29	28	24
7	67	37	46	251	21	50	52	48	37	267	267	265	259	23	24	22	15
8	94	77	65	326	37	79	80	76	65	349	350	341	327	40	40	38	31
9	71	64	60	296	30	65	65	65	60	310	307	311	315	31	31	30	28
10	55	28	28	221	14	37	39	36	28	230	230	229	237	15	17	14	10
11	64	28	50	238	18	47	50	45	28	260	261	256	237	22	23	21	10
12	95	97	98	400	49	97	97	96	95	400	400	400	400	49	49	49	48
13	43	21	15	200	10	26	29	25	15	200	200	200	205	9	10	9	6
14	66	53	53	271	25	57	58	57	53	288	285	289	298	27	27	26	25
15	74	60	62	286	28	65	66	65	60	311	309	311	315	31	31	30	28
16	36	13	32	200	10	27	29	26	13	202	200	203	200	10	10	9	5
17	64	51	39	256	22	51	52	50	39	271	269	270	263	24	24	23	17
18	77	74	48	301	31	66	68	64	48	314	314	308	285	31	32	30	22
19	75	51	58	271	25	61	62	60	51	300	298	298	293	29	29	28	24
20	47	18	35	213	12	33	35	32	18	220	219	219	212	13	14	12	7
21	93	88	90	379	46	90	90	90	88	382	381	382	383	46	46	46	45
22	61	41	47	251	21	50	50	49	41	266	264	266	268	23	23	22	18
23	82	77	70	338	40	76	76	76	70	342	341	342	339	38	38	38	34
24	88	75	75	344	41	79	80	78	75	351	350	349	351	40	40	39	37
25	84	67	63	305	32	71	72	70	63	328	327	325	322	35	35	34	30
26	63	44	28	238	18	45	47	43	28	253	254	250	237	21	22	19	10
27	92	95	94	393	48	94	94	94	92	391	391	392	393	47	47	47	47
28	58	56	51	271	25	55	55	55	51	282	277	283	293	26	26	25	24
29	62	48	57	266	24	56	56	55	48	283	280	284	285	26	26	26	22
30	70	68	64	305	32	67	67	67	64	317	314	318	324	32	32	32	30
31	55	43	48	251	21	49	49	48	43	264	259	265	273	22	22	22	19
32	69	48	41	256	22	53	54	51	41	275	274	273	268	24	25	24	18
33	80	84	71	358	43	78	79	78	71	348	347	347	341	39	40	39	35
34	90	78	76	351	42	81	82	80	76	356	355	355	354	41	41	41	38

35	57	26	41	234	17	41	43	40	26	243	243	241	232	18	19	17	9
36	74	74	72	326	37	73	73	73	72	334	331	335	344	36	36	36	35
37	87	68	83	330	38	79	80	78	68	351	350	347	334	40	40	39	33
38	74	33	35	238	18	47	51	44	33	260	265	253	249	22	23	20	13
39	90	78	81	358	43	83	83	82	78	361	360	360	359	42	42	42	39
40	62	45	71	271	25	59	60	58	45	294	293	292	278	28	28	27	21
41	71	51	49	266	24	57	58	56	49	287	286	286	288	27	27	26	23
42	39	32	27	226	15	33	33	32	27	218	212	220	234	12	13	12	10
43	69	48	32	246	20	50	52	47	32	266	268	262	246	23	24	22	12
44	65	64	43	271	25	57	58	56	43	288	287	287	273	27	27	26	19
45	80	63	58	291	29	67	68	66	58	316	314	313	310	32	32	31	27
46	74	55	59	271	25	63	63	62	55	303	301	302	302	29	30	29	26
47	83	76	71	338	40	77	77	76	71	343	342	343	341	38	38	38	35
48	44	20	44	226	15	36	38	35	20	227	226	227	217	15	16	14	7