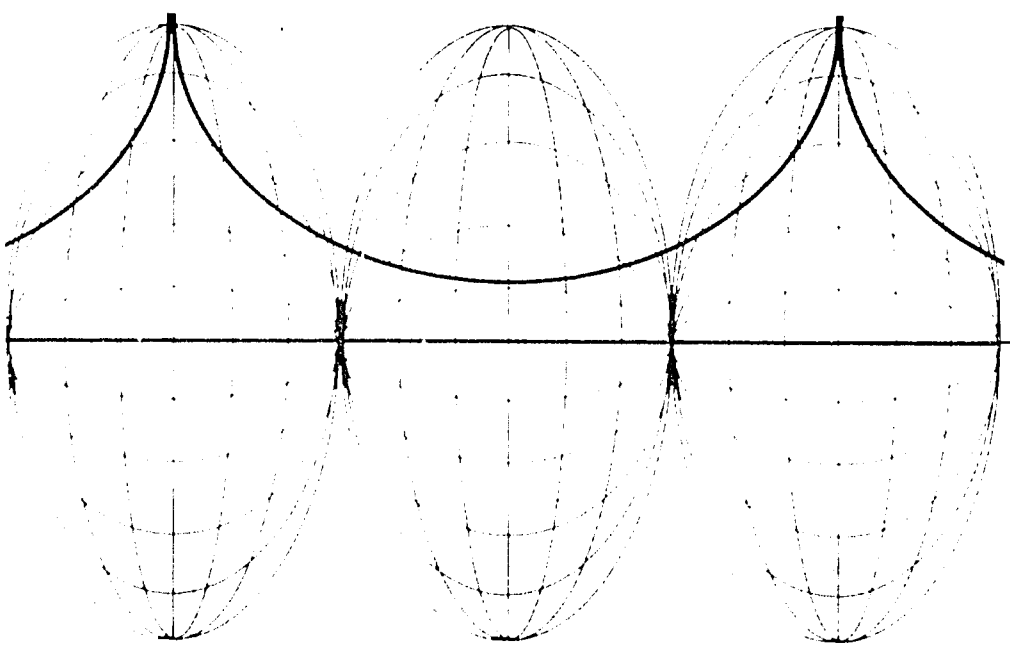
PN-ABE-179

A Workshop on Educational Policy Analysis and Planning

August 1988



B • R • I • D • G • E • S

Basic Research and Implementation in DevelopinG Education Systems

The Basic Research and Implementation in DevelopinG Education Systems Project (BRIDGES) is directed by the Harvard Institute for International Development and the Harvard Graduate School of Education, under Cooperative Agreement No. DDP-5824-A-5076 with the Office of Education, Bureau for Science and Technology, United States Agency for International Development. Also participating in the Project are the Institute for International Research, Michigan State University, the Research Triangle Institute, and Texas Southern University.

The BRIDGES Project includes educators, researchers, planners, and policy makers committed to improving opportunity and quality in Third World Schools. The goal of their collaborative effort is to identify policy options that will increase children's access to schooling, reduce the frequency of early school leaving and repetition, improve the amount and quality of what is learned, and optimize the use of fiscal and educational resources.

All publications generated by project activities are held in the BRIDGES Publications Office at Harvard University. Requests for project documents should be addressed to:

Project BRIDGES
Gutman 442
Harvard University
Cambridge, MA 02138
USA
Telephone: 617/495-9720
Telex: 275276
Fax: 617-495-0527

United States Agency for International Development
Bureau for Science and Technology
Office of Education
Contract No. DPE-5824-A-00-5076-00

A Workshop on Educational Policy Analysis and Planning

August 1988

Project BRIDGES
Basic **R**esearch and **I**mplementation in **D**evelopin**G** **E**ducation **S**ystems

BRIDGES
A Workshop on Educational Policy Analysis and Planning

Table of Contents

Section I: BRIDGES Workshop Overview

Objectives of the Workshop
Schedule of events for each week

Section II: Educational Planning by Russell G. Davis

Chapter 1: The Foundations of Planning - Goals and Objectives
Chapter 2: General Background on Planning Models
Chapter 3: Population Forecasting
Chapter 4: Enrollment Forecasting
Chapter 5: Manpower Planning
Chapter 6: Management Information Systems

Assignments 1-5 (Yellow Paper)

Section III: Educational Policy Analysis by William Cummings

Chapter 1: The Variety of Problems for Educational Research
Chapter 2: Clarifying the Problem and Preparing the Research Design
Chapter 3: Considerations for Evaluating and Editing Research Reports
Chapter 4: Conducting and Reporting Multivariate Analysis

Code Book for Data Set
Assignments 6-9 (Yellow Paper)

Section IV: System for Tracking Educational Progress (STEP) by Luis Crouch and Scott Moreland

Chapter 1: Introduction to the STEP System
Chapter 2: Using STEP
Chapter 3: Configuring STEP to a New Application
Chapter 4: Technical Description of the Model Methodology

Section V: Microcomputer Information

Diagram of a Microcomputer System
Introduction/Review of Basic LOTUS 1-2-3 Commands
Getting Started with SYSTAT

Section I: BRIDGES Workshop Overview

Workshop Objectives

The objectives of the BRIDGES Workshop on Educational Policy Analysis and Planning are:

* To contribute to participants' understanding of current concepts and methodologies of educational policy analysis and planning.

* To encourage the use of data routinely collected by ministries of education and known analytical techniques for policy analysis.

* To aid participants with the development of systematic outlines of information requirements for policy analysis and planning in their countries.

* To provide participants with knowledge about the structure of the hands-on experience with microcomputer software packages useful for policy analysis and planning.

* To train participants in the use of the System for Tracking Educational Progress (STEP), a microcomputer software package designed by BRIDGES to aid in projecting basic educational indicators, calculating selected educational efficiency measures, and estimating resource requirements.

BRIDGES Workshop on
Educational Policy Analysis and Planning

Schedule

Week 1

Day 1

Morning		Opening Exercises
		Welcome and Introduction to the Workshop
		Introduction of the Workshop Staff
		Introduction of the Participants
10:30-12:00		Orientation
Afternoon 1:30-3:15		Lecture, Demonstration Topic: Introduction to Educational Planning
	***	**Reading Assignment due:** "Educational Planning", Chapter 1
3:30-4:30		Lecture, Demonstration Topic: Introduction to Microcomputers
4:30-5:00		Tour of Library

Day 2

Morning 9:00-10:30		Lecture, Demonstration Topic: Educational Modeling
	***	**Reading Assignment due:** "Educational Planning", Chapter 2
10:45-12:00		Lecture, Demonstration Exercise: Spreadsheet Software for Modeling
Afternoon 1:30-3:00		Demonstration Hands-On Introduction/Review: LOTUS 1-2-3
3:15-5:00		Hands-On Microcomputer Activities

Day 3

Morning 9:00-10:30		Lecture, Demonstration Topic: Forecasting Models: Population/Demographic
	***	**Reading Assignment due:** "Educational Planning", Chapter 3
10:45-12:00		Demonstration Exercise: The POPEX Model
Afternoon 1:30-5:00		Hands-on Microcomputer Activities Work on: Assignment #2: Evaluation of a software model or package
Evening 7:00-9:00		Microcomputer lab open for individual work.

Day 4

Morning 9:00-10:30		Lecture, demonstration Topic: Forecasting Models: Enrollment
	***	**Reading Assignment due:** "Educational Planning", Chapter 4
	***	Assignment due: #1-Evaluation of an Educational Plan
10:45-12:00		Lecture, demonstration Topic: Enrollment Forecasting Continued
Afternoon 1:30-5:00		Hands-on Microcomputer Activities Work on: Assignment #3: Enrollment Flow Exercise Continue work on other assignments
Evening 7:00-9:00		Microcomputer lab open for individual work

Day 5

Morning
9:00-10:30 — Lecture, demonstration
Topic: Manpower Planning

10:45-12:00 — Lecture, demonstration
Topic: Manpower Planning Continued

*** **Reading Assignment due:**
"Educational Planning", Chapter 5

Afternoon
1:30-3:00 — Lecture, demonstration
Topic: Manpower Planning Continued

3:15-5:00 — Lecture, demonstration
Topic: Time series analysis
Exercise: Using Smart Forecasts II

Week 2
Day 6

Morning
9:00-10:30 — Lecture, demonstration
Topic: Financial Models for Planning:
IFPS, EDFISIMO, and other LOTUS models

10:45-12:00 — Lecture, demonstration
Topic: Financial Models for Planning

Afternoon
1:30-3:00 — Lecture, demonstration
Topic: Financial Models for Planning
Continued

3:15-5:00 — Demonstration: Financial Models

Evening
7:00-9:00 — Microcomputer Lab open for individual work.

Day 7

Morning
9:00-10:30 — Lecture, demonstration
Topic: Management Information Systems
Exercise: Relational Databases

*** Reading Assignment due:
"Educational Planning", Chapter 6

*** **Assignment due:** #3-Enrollment Flow Exercise

10:45-12:00	Lecture, demonstration Topic: Management Information Systems Development: Egypt
Afternoon 1:30-5:00	Lecture, demonstration Topic: Introduction to R:Base System

Day 8

Morning 9:00-10:30	Discussion Topic: Management Information Systems for Education
10:45-12:00	Lecture, demonstration Topic: R:Base Continued
Afternoon 1:30-3:00	Lecture, demonstration Topic: R:Base Continued
3:15-5:00	Hands-on Microcomputer Activities Continue work on all assignments and exercises
Evening 7:00-9:00	Microcomputer Lab open for individual work.

Day 9

Morning 9:00-10:30	Lecture, demonstration Topic: Introduction to Sri Lanka Database
10:45-12:00	Lecture, demonstration Topic: Relating Databases with Statistical Packages
Afternoon 1:30-5:00	Hands-on Microcomputer Activities Complete all assignments and exercises
Evening 7:00-9:00	Microcomputer Lab open for individual work.

Day 10

Morning
9:00-10:30 Small Group Presentations

*** **Assignment due:** #2-Evaluation of a Software Package or Model

*** **Assignment due:** #4-Creating a Personnel Database

*** **Assignment due:** #5-Synthesis and Application of Weeks 1 and 2

10:45-12:00 Lecture, demonstration
Topic: Summary and Review of Educational Planning

Afternoon
1:30-3:00 Lecture, discussion
Topic: Introduction to Statistical Packages and Comprehensive Systems
Topic: Transition from Weeks 1 & 2 to Weeks 3 & 4

3:15-5:00 Lecture, discussion
Transition Continued

Week 3

Day 11

Morning
9:00-10:30 Lecture, demonstration
Topic: Introduction to Policy Analysis

*** **Reading Assignment due:**
"Educational Policy Analysis", Chapters 1 & 2

10:45-12:00 Lecture, demonstration
Topic: Policy Analysis Continued

Afternoon
1:30-3:00 Small Group Activities
Topics: Planning policy studies, Merging files, Creating variables

3:00-3:15 Break

3:15-5:00 Small Group Activities Continued
Work on Assignment #6

Evening
7:00-9:00 Microcomputer Lab open for individual work.

Day 12

Morning
9:00-10:30 — Lecture, demonstration
Topic: Policy Analysis Continued

*** **Assignment due:** #6-Identifying and Creating Variables

10:45-12:00 — Lecture, discussion
Topics: Univariate analysis, Cleaning data, Examining relations between variables, Preparing tables

Afternoon
1:30-5:00 — Small Group Activities
Work on Assignment #7

Day 13

Morning
9:00-10:30 — Lecture, demonstration
Topic: General Principles on Preparing Reports

*** **Reading Assignment due:**
"Educational Policy Analysis", Chapter 3

*** **Assignment due:** #7-Univariate Analysis and Cleaning Data

10:45-12:00 — Small Group Activities
Work on Assignment #8

Afternoon
1:30-3:00 — Lecture, demonstration
Topic: Introduction to SYSTAT

3:15-5:00 — Small Group Activities
Work on SYSTAT and Assignment #8

Evening
7:00-9:00 — Microcomputer Lab open for individual work.

Day 14

Morning
9:00-10:30 — Lecture, demonstration
Topic: Introduction to Multivariate Analysis

*** **Reading Assignment due:**
"Educational Policy Analysis", Chapter 4

*** **Assignment due:** #8-Using a Statistical Package

10:45-12:00 — Lecture, demonstration
Topic: Multivariate Analysis Continued

Afternoon
1:30-5:00 — Lecture, demonstration
Topic: Introduction to Graphics

Evening
7:00-9:00 — Microcomputer Lab open for individual work.

Day 15

Morning
9:00-10:30 — Small Group Activities
Work on Assignment #9

10:45-12:00 — Small Group Activities
Work on Assignment #9

Afternoon
1:30-3:00 — Small Group Presentations
Presentation of Assignment #9

*** **Assignment due:** #9-Conducting a Policy Study

3:15-5:00 — Small Group Presentations Continued

Week 4

Day 16

Morning
9:00-10:30 — Lecture, demonstration
Topic: Introduction to the System for Tracking Educational progress (STEP) and HOST

10:45-12:00 — Lecture, demonstration
Topic: STEP Continued
Use of the ESTIMATE Module

Afternoon
1:30-3:00 — Lecture, demonstration
Topics: Use of the PROJRATE and EDPROJ Modules

3:15-5:00 — Small Group Activities
Hands-On Practice with STEP

Day 17

Morning 9:00-10:30	Lecture, demonstration Topic: STEP Continued Use of the COHORT Module
10:45-12:00	Small Group Activities Practice with the COHORT Module
Afternoon 1:30-5:00	Small Group Activities Hands-On Practice with STEP

Day 18

Morning 9:00-10:30	Lecture, demonstration Topic: STEP Continued Use of the POP Module Demonstration of Cost Module
10:45-12:00	Lecture, demonstration Topic: STEP Continued
Afternoon 1:30-3:00	Lecture, demonstration Topic: Introduction to BRIDGES Educational Planning and Policy Game Use of EDUCATE, POPULATE, and ECON Modules
3:15-5:00	Small Group Activities Hands-On Practice with STEP Work on Assignment #10

Day 19

Morning 9:00-10:30	Lecture, demonstration Topic: BRIDGES Educational Plannning and Policy Game Continued
10:45-12:00	Small Group Activities Topic: BRIDGES Game
Afternoon 1:30-5:00	Small Group Activities Topic: BRIDGES Game Practice with STEP

Day 20

Morning

9:00-10:30		Lecture, discussion Topic: Application of STEP
	***	**Assignment due:** #10-Application of STEP to Data from Home Countries
10:45-12:00		Small Group Discussion

Afternoon

1:30-3:00	Workshop Wrap-Up
3:15-5:00	Small Group Discussion Implementing Workshop principles and practices at home: Issues and Suggestions

Section II: Educational Planning

EDUCATIONAL PLANNING

by

Russell G. Davis

Table of Contents

Chapter 1	**The Foundations of Planning: Goals and Objectives**
	Assignment #1: Evaluation of an Educational Plan.
Chapter 2	**General Background on Planning Models**
	Exercise: Spreadsheet Software for Modeling.
Chapter 3	**Population Forecasting**
	Exercise: The POPEX Model.
	Assignment #2: Evaluation of a Software Package or Model.
Chapter 4	**Enrollment Forecasting**
	Assignment #3: Enrollment Flow Analysis Exercise.
Chapter 5	**Manpower Planning**
	Exercise: Using Smart Forecasts II.
Chapter 6	Management Information Systems
	Exercise: Relational Databases.
	Assignment #4: Creating a Personnel Database.
	Assignment #5: Synthesis and Application of

Chapter 1

THE FOUNDATIONS OF PLANNING: GOALS AND OBJECTIVES

1.0 WHAT DO PLANNERS DO?

Within the educational system there are several types of personnel. Some work directly with students and schools. Teachers help people acquire necessary knowledge, values, attitudes, and skills. Supervisors and specialists work with teachers to support them in providing quality education. Officials manage the combined effort of schooling. They manage the system well when (1) students learn most, best, with the greatest amount of satisfaction, and (2) when the least amount of time and money is spent.

Educational planners do not work in schools, but they help in the management of schools by describing the fit between what students are presently learning and the actual learning needs and goals of the system. Planners not only describe this fit (or lack of fit), but they also recommend how it can be improved over a period of time. They also assist in determining what the future learning needs and goals should be.

Because planners deal with broad objectives over a long time horizon, they often use the term goals instead of the word objectives. There is often no clear difference between the two terms. However, goals are usually considered broad statements of the ends, purposes, or outcomes of education. Goals are referenced to ends, purposes or outcomes that are beyond the school. Goals state the wider social purpose objectives. Objectives are more narrow statements describing what people should know or be able to do as a result of learning. Here are two examples:

"The learner will be able to recite the multiplication table".

This is an educational objective. It describes a skill learners should have at the end of an instructional unit.

"75% of all secondary students will master the arithmetical calculations necessary to work as engineers".

1.1 LEVELS OF PLANNING

1. Systems Level Planning

Planning at the systems level generally begins from goals, where the system is a set of individual schools that share common goals, organizational direction, and base of support. Planning may also begin with needs but, because planners in national school systems work in a central ministry which may be far removed from direct contact with learners and their needs, it is more common for systems planning to be based on goals and objectives derived from needs.

2. Institutional Level Planning

Institutional level planning is also based on goals and objectives set for a large institution at the secondary or higher education level. The institution can be a system in itself, as in the case of a university with multiple schools, campuses, institutes, or services.

3. Program or Project Level Planning

Within institutions there are programs connected in fields of study. The programs have:

- o <u>objectives</u> that specify the learning outcome or what the program is designed to accomplish,
- o <u>learning strategies</u> that map out the instructional activities (the teaching/learning events), and
- o <u>resources</u> to support the activity.

A program exists within an institution; a project is free standing or independent of an institution, but otherwise the planning for the two is similar. From programs, planning can go down to <u>course level</u>, which means the planning of a <u>learning unit</u>, or even a <u>micro teaching</u> segment. The material which follows will be restricted to <u>systems planning</u>, but planners must sometimes work at the institution or program level.

1.2 THE RATIONALE FOR PLANNING

Educational planning sets clear directions for how schools are to attain their objectives. This is important because there are many different ways to reach objectives, but time and money are always limited. In mapping out strategies and programs for attaining objectives, planners seek a balance between effectiveness and efficiency:

Effectiveness means that the objectives will be reached (or attained to some satisfactory level).

Efficiency means that the objectives will be attained at the least cost in time and money.

At the same time, planners must also be certain that the goals of the schools are being met. Ideally, attaining the objectives should lead to wider economic benefits in the form of increased production or wider political benefits in the form of national unity. When that is true, education is also reaching the goals set for schools in the society.

Making sure that the objectives agree with the goals is the most difficult task in educational planning. Because of this, planning works best when it:

1. States clear goals.
2. States clear objectives.
3. Maps out strategies or alternative ways to reach the stated goals and objectives.
4. Helps to choose the means, policies, and programs to attain the objectives at the lowest cost in time and money.

Notice that the sequence above shows that goals should be established before objectives. This makes it more likely that the objectives will be in agreement with stated goals. When the objectives are in agreement with the goals, attaining an objective will automatically help in attaining a goal. (In reality, of course, planners sometimes must deal with cases where no goals have been stated. In such cases, planning must start at the opposite end with program or even unit objectives. "Rationality" in planning and "reality" do not always agree!)

Figure 1.1 on the next page is a more complete outline of the idealized stages of planning. Study it carefully before continuing.

Figure 1.1: Outline of the Idealized Stages of Planning at Education Systems Level

I. Establishment of Goals

Broad statements of social, political, and economic ends and purposes. Establishing goals will take a great deal of discussion and debate. Priorities must be set; needs must be considered in relation to values.

II. Consideration of Broad Strategies for Attaining Goals

1. **Population:** Who are the target groups that will be affected by the goals?
2. **Activities, Assets, Services, Processes:** What resources will be required to reach the goals?
3. **Human Resource Development:** How can education and training assist in reaching the goals?

III. Consideration of Alternative Means for Attaining Goals

1. Is Human Resource Development the most effective means for reaching the goals, or
2. Should the goals be attained through the establishment of economic, political, or social policies and/or programs?

IV. Consideration of Alternative Policies to Support the Means

Once the appropriate means are determined, policies may need to be written into law or legislation (or into institutional or program regulations) to make further development possible and practical. Existing policies must be assessed to see that they do not conflict with goals and means.

V. Consideration of Alternative Programs for Attaining Goals

Assuming that Human Resource Development is the appropriate means and that relevant policies are in place, what kinds of educational or training programs should be developed and how should they be implemented? For each program, consider:

- **Program Objectives** (or Institutional Objectives)
- **Program Activities**
- **Program Modalities** (Organization, Delivery Modes, ...)
- **Needed Resources** (Teachers, Materials, Facilities, ...)
- **Costs and Budgets**

VI. Evaluation of Effects

Effects should be traced back through the program objectives and ultimately to the original goals. Evaluation will include assessment, evaluation, cost/benefit, cost/effectiveness analysis. **Feedback** should be recirculated throughout the entire system and may lead to revision at any level if required.

1.3 AN EXAMPLE FROM EGYPT

In order to see how the elements shown in Figure 1.1 appear in an actual national planning document, an example will be presented. The source of the example is the Egyptian Five Year Plan, 1987/88-1991/92.

GOAL:

"Since Egypt has a very limited supply of natural resources, it has to maximize its investments in its human resources through education and training in order to satisfy the labour requirements of its economic and social development plans..."

This is an almost perfect example of an educational goal. Note that:

- It begins with a need or problem--the lack of natural resources. This need/problem gives rise to a goal response, the necessity for a greater effort to educate human resources.

- The educational goal of developing required manpower through education and training is referenced to the need previously stated and to an outcome in the wider economy.

- The statement provides a criterion for monitoring the attainment of the goal--in this case the fit between training output and manpower requirements. This is rare, but it provides a goal framework within which planners can work.

Corresponding to this goal would be thousands of objectives stated for the institutions (schools, centers), programs (fields of learning within institutions), and special training projects that make up the national education/training system in Egypt. Planning must fit institutional, program, and project level objectives to national systems level goals. (Eventually, even micro unit objectives will be tied to the goal.) Notice how closely the three objectives on the next page are linked to the goal statement.

1. Goal

Since Egypt has a very limited supply of natural resources, it has to maximize its investments in its human resources through education and training in order to satisfy the labour requirements of its economic and social development plans..."

2. Institutional Level Training Objective

The Mubarak Center will produce the skilled workers and trained technicians in agricultural machinery that are required to meet the manpower needs of agricultural development in lower Egypt.

3. Program Level Learning Objective

Upon completion of this course, students will understand the basic principles of diesel mechanics and be able to: diagnose causes of failure in a diesel engine, take down the engine, fix it, reassemble it, test it, and certify it as repaired.

4. Micro Unit Learning Objective

After completing this unit, students will be able to identify the names of the main parts or components of a pump or tractor engine.

The point of the example is that planning should help educators relate goals at the systems level to objectives at institutional, program, course, or even unit level. In cases where systems goals have not been specified, planners may need to help build up to systems goals from lower level objectives. Without goals and objectives that are (1) clearly understood and (2) closely linked together, there is no rational way for determining how effective or efficient a system is operating.

Once the goals and objectives are clear, specific targets (descriptions of gaps between actual learning in the schools and the needs and goals of the system) can be specified for accomplishment. For example, suppose a goal for a system states that "80% of the age eligible children will enter the first level of the system". Research on admissions then finds that only 68% of age eligible children are being admitted. Obviously, there is a gap between what is specified in the goal and reality. Planners may at this point set a target such as:

> "Over the next five years, the intake rate of age eligible students entering the first level of the system will be raised from 68% to 80%."

Now the task becomes determining how to meet the target, thereby accomplishing the goals. It is important to realize that the planner does not yet begin to state program objectives. Rather, as shown on Figure 1.1, there are several intermediate stages. These stages are particularly important because goals are often very general statements (and unfortunately sometimes vague) and objectives are specific (and sometimes overly precise). The planner must bridge the gap between the two. This is accomplished with strategies and policies. Strategies and policies are discussed in the next section.

1.4 LINKING GOALS TO OBJECTIVES WITH STRATEGIES AND POLICIES

Strategies lay out in broad terms the alternative ways, the alternative means for attaining goals. Strategies specify how means conflict and interfere, or how means support and help each other. Since all means cannot be pursued at one and the same time, strategies specify the best sequences, the priorities for choosing, deciding on, and implementing the means.

Policies specify (1) how strategies and programs are to be carried out, (2) who will participate in the programs and how, and (3) what the programs will provide and under what conditions. Policies may be written into law or legislation, or in institutional or program regulations.

Rather than discussing abstract definitions, another example will be presented. The example will trace the path between the goals stated in a national plan and the actual implementation of instruction, noting the strategies and policies that make the implementation possible.

The set of goals, policies, and objectives that follows comes from the Second Development Plan of Saudi Arabia (1975-1980). The plan starts with the fundamental goals of society:

1. Goal Statement

Maintain the religious and moral values of Islam
Insure national defense and security
Insure economic development through high growth
Reduce dependence on oil exports and preserve reserves
Foster social stability along with social change
Develop human resources through education/training and health.

2. Strategies, Objectives, Policies

Within the broad strategy of Human Resource Development, objectives, policies, programs, and projects would be developed to cover four major areas:

- o Manpower development
- o Training (Public administration and vocational training)
- o Labor law, services, security
- o Education

Then, each of these four areas must set general objectives for all of its' units. For example, Education, would set general objectives for its' four main branches:

- o Boys' Education
- o Girls' Education
- o Higher Education
- o Religious Education

Within each of these branches, policies and program objectives are set. As an example, Boys' Education will be described in more detail on the next page.

3. Boys' Education: Objectives, Policies, Programs

The general policy objective for Boys' Education is to provide quality education through secondary school level for all boys. Some of the programs for Boys' Education include:

- o Pre-elementary
- o Elementary
- o Intermediate
- o Secondary
- o Technical Education
- o Teacher Training

In the actual Saudi plan it is clear that some of the program branches (such as Intermediate Education) established policies and objectives and then set targets:

4. Intermediate Education: Objectives, Policies, Targets

Objective: Provide broad base of quality education to prepare for further education and training

Policy: Assure that opportunities for intermediate education are provided in rural areas.

Target: Enroll all, graduate 95 percent of entrants by end of Second Development Plan (1980).

Other branches of Boys' Education in the Saudi plan skipped writing any objectives and policies and went directly to quantitative targets:

5. Technical Education: Target Only

Target: Expand, in order to graduate 1600 students by 1980.

Plans are weakened when they go directly from broad goals to enrollment targets without intervening links to policies and objectives. The whole purpose of the activity is left out. It is even weaker to go directly to targets stated in terms of inputs. The Saudi plan did this in the

Higher Education section. There was little review of objectives or policies, and the plan went directly to the statement of activity input targets:

"Complete construction work at ad-Dir'iyah campus".

"Ensure support personnel are provided."

These statements were not linked back to program objectives, policies, strategies, or the general goals for human resource development. Thus, there was no check for:

- o Whether the inputs were related to program objectives
- o Whether program objectives were consistent with policies and strategies
- o Whether input activities, program objectives, policies and strategies were related to the general manpower, economic growth, and human resource development goals stated earlier.

Insuring that goals, objectives, and activities do not conflict and that programs are related to goals are the main tasks (and challenges) of planning. The example above showed some effective and some ineffective planning. Some of the educational planning showed strong links to national development goals, to educational sector strategies and policies, and to objective and activity targets. Some did not.

Summary: General Problems in Educational Planning

The two most general failings of educational planning are:

A. To offer vague and general goals without establishing any path or relationship to later educational objectives, targets, or lists of activities.

B. To provide long and detailed lists of targets, activities, and inputs for activities, without establishing any linkage to objectives and goals of education. Often the plan is simply a laundry list of actions and inputs, with no mention of the goals or purposes.

The first kind of planning sets impressive goals, but without providing any directions for reaching them. The second kind of planning provides specific details on the journey, but never describes the destination. Planners can avoid the extremes of the problem if they keep in mind that goals, however fine, must be connected to strategies, policies, objectives, and targets.

The following chapters assume that the essential goals, strategies, and objectives for a system have been set. When this is true, the planner must then

(1) Analyze the gaps between stated goals/objectives and actual performance,

(2) Set targets to lead the system toward eventually closing these gaps, and

(3) Forecast how the targets are to be attained over time.

To accomplish these tasks, planners use a variety of techniques and models. A model is a simplified picture of reality. There are two main types of models:

Heuristic models help to clarify goals, develop goal structures, and relate goals to objectives.

Algorithmic models are used to analyze and forecast variables.

The next chapter will focus on algorithmic models, but it is important to note that planners must deal with both kinds of models. Many planners (and their critics) fail to realize that all of the tasks of planning are not done with algorithmic models and methods. Planning is based on goals--and goals are dealt with by using heuristic methods.

APPENDIX A

PLANNING POLICY FROM THE PERSPECTIVE OF THE DECADE AHEAD:

A FUTURES-CREATING PARADIGM

(Pages 14-19)

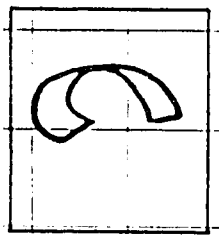
[Source: American Association of State Colleges and Universities. Resource Center for Planned Change, 1978.]

stage 1

societal trend areas

The context of societal trends profoundly affects the mission of the college or university. The areas within society from which trends can be specified and traced will be called societal trend areas. These societal trend areas can be identified and then component trends traced with reasonable accuracy and a sense of confidence for a decade or so; their effects can be anticipated.

The following set of such societal trend areas has been developed for the purpose of illustration.



- population
- government
- global affairs
- environment
- energy
- economy
- science and technology
- human settlements
- work
- lifestyle
- women
- participation

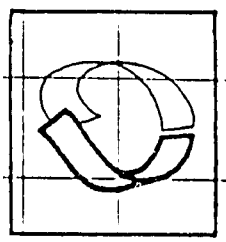
The set of societal trend areas is not exhaustive. Other areas may be of equal or greater importance and perhaps more germane to any one institution's needs. Stage 1, then, calls for the institution to identify both national and regional trend areas *it believes* will impact policy planning in its management of change.

stage 2

societal value shifts

Similarly, the context of values—general societal, regional, and local, and those within each academic community—will influence institutions in their management of change. Values, too, can be defined, selected, described, and analyzed. Such activity will bring about an informed anticipation of value shifts that can be incorporated into institutional policy making.

As in the case of the trend areas, 12 apparent value shifts are described and labeled for the purpose of illustration.



- change
- freedom
- equality
- leisure
- foresight
- pluralism
- localism
- responsibility
- knowledge
- quality
- goals
- interdependence

Of course the 12 value shifts do not exhaust the possibilities. There are a multitude of other societal values; there are also the special regional or local value systems and the specific institutional values that must be taken into consideration. Indeed, any one of those—or a combination of them—may modify and further influence the broader values.

Stage 2 of the paradigm, then, calls for the institution to identify *for itself* those national, regional, and institutional value shifts it believes will impact its planning.

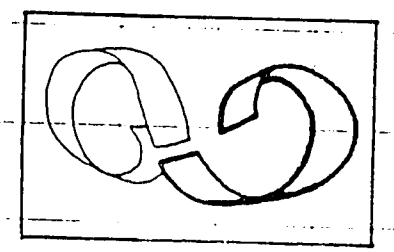
Parenthetically, the placement of trends before values in this guide is dictated by no weightier consideration than simple logistics. The roles of trends and value shifts are of equal importance in the consideration of the institution's future.

stage 3

a matrix for trend areas...

Once the societal trend areas have been described, it becomes possible to project through the utilization of the policy-making matrix the effects these trends are likely to cause on selected sectors of the institution. The sectors, identified for illustration, are: curricula, faculty, students, research and development, public service, internal resource allocation, facilities, external resources, administration. The projections of the effects on these sectors will vary. Both desirable and undesirable effects may emerge.

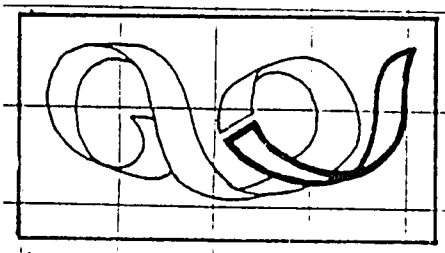
Stage 3, then, calls for the construction of a policy-making matrix for the trend areas comprising the results of a careful analysis of these trends and their likely impacts on the various sectors of the institution over the next 10 years.



stage 4

...and a matrix for the value shifts

Stage 4 calls for the construction of a separate policy-making matrix for tracing the impacts of each particular value shift upon each sector of the institution.

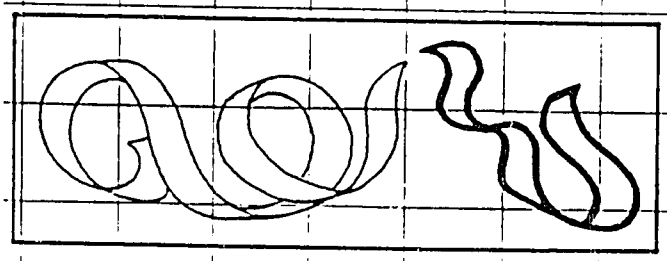


Both policy making matrices (Stages 3 and 4) present cells of dynamic information that will furnish later stages of the planning process with the impact data and the images of the future needed to proceed.

stage 5

institutional objectives

The development of policies to guide the institution into and beyond the next decade can now be initiated. When the matrices have been completed and the desirable and undesirable effects of the anticipated trends and shifts in the value structure have been traced, the planners are ready to formulate institutional objectives. (The reader should recall that policies contain objectives and the means for their achievement.) There is no limit to the number of objectives to be developed; all are to be included in a comprehensive set of objective statements.



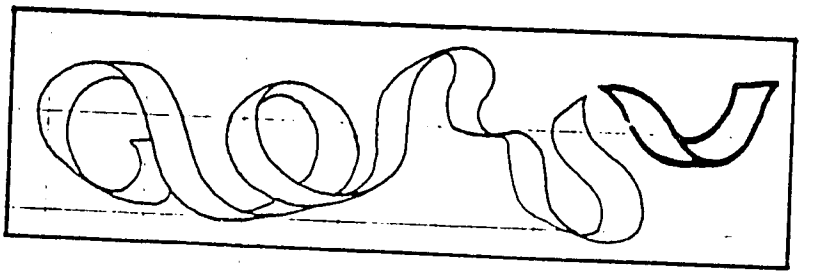
Stage 5, then, calls for the planners to develop institutional objectives. Once they have been stated, objectives must be surrounded by data. Thus, the planners must further describe each in relation to the institution's mission, the impacts of societal trends and value shifts, and specific indicators of the objective achievement.

stage 6

testing compatibility

Planners can anticipate possible difficulties ahead by testing the compatibility of their objectives

micro 4



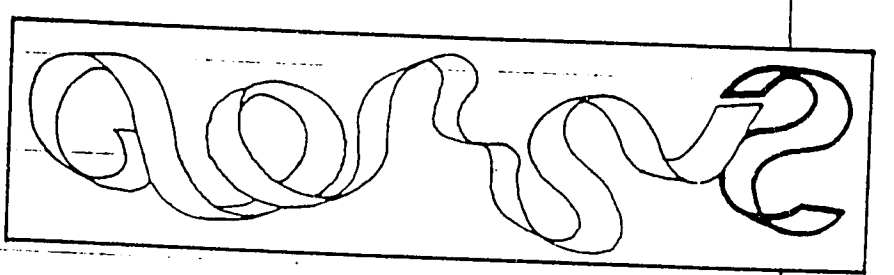
To this end, compatibility must be ascertained between objectives and (1) the institution's current mission statement, (2) shifts in the value system, and (3) each of the other objectives formulated by the planners. Evidence of incompatibility can be considered in light of additional data to be generated in later planning stages in order to make judgments about pursuing, modifying, or abandoning proposed policies.

stage 7

a futures scenario

Stage 7 is a creative step in the paradigm. It calls for "fleshing out" the objective statements and portraying, in graphic terms, the cumulative impact of an entire set of events. It does so by exploring, in narrative form, (1) the most optimistic and (2) the most pessimistic combinations of events that are likely to happen. The first scenario assumes the achievement of an objective and how this realization will affect future conditions. The second imagines the future of the institution should the objective not be undertaken for whatever reason.

Compelling information regarding the importance of each objective will result from the description of projected conditions and the invention of results that may or may not occur.

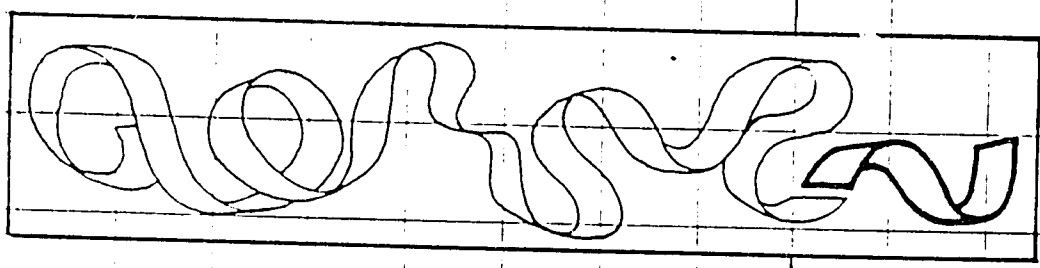


stage 8

a history of alternative futures

Thus far, exploration of the contextual areas of societal trends

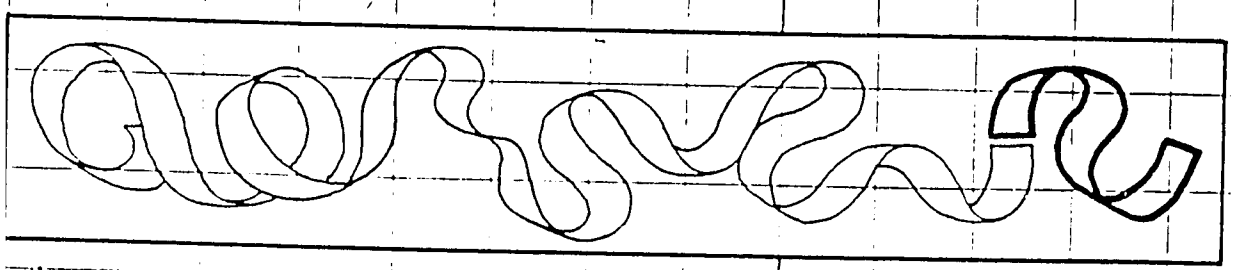
and value shifts have yielded two data banks of pertinent information, the formulation of institutional objectives, and a greater understanding of the importance of all objectives through scenario building. In order to move from consideration of desirable objectives toward the further development of institutional policy, the planners now must think about strategies and tactics. The paradigm calls for Stage 8 to build a futures history. What this stage requires is planning from the future, that is, to set a specific year for the objective's realization, say 1989, and then plan back from it stating what events, strategies, and finally tactics will have had to take place in order to bring about that objective. This is the essence of planning *from* the future *for* the future because, with Aristotle, the ideas must now be created (potentially) and transformed into events (actuality).



stage 9

foresight

This stage explores the effects, the consequences, and the results of the policies that have been proposed. A specific activity integral to the planning paradigm, this assessment concerns the examination of the final condition brought about by the implementation of policy.



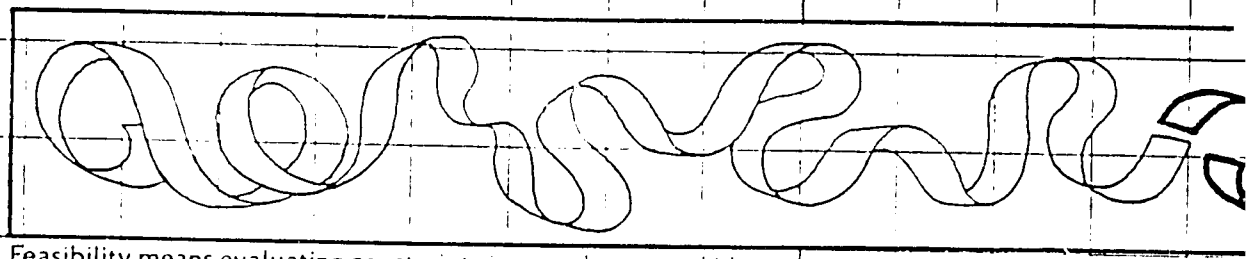
urther, an analysis can be made of the impacts, not only of the bjectives, but also of the strategies and tactics on the sectors of ie institution, the institution's purpose, and the community in hich the institution functions.

microby

stage 10

feasibility

Of course, under the institution's ability to implement policy lies the question of feasibility. The exploration of this critical issue is taken up at the end of the process in order to allow maximum margin to the creative potential of the Futures-Creating Paradigm. Simply, the flow of the process should not be interrupted; it must be permitted to open up and go as far as the planners are able to take it.



Feasibility means evaluating constraints imposed upon and identified as a result of planning. National, sometimes universal, certainly regional and institutional perspectives must be considered. Clearly, some or all of an institution's characteristics and requirements which color that perspective may present inhibiting factors that must be addressed. Stage 10, then, recognizes the relevance of feasibility in the execution of the policy-planning process; it asks the institution to determine the feasibility of the proposed policies as weighed against societal and institutional constraints.

Chapter 2

GENERAL BACKGROUND ON PLANNING MODELS

2.0 AN INTRODUCTION TO MODELS AND VARIABLES

A model is built to show reality in simple form by words, pictures, or algebraic symbols. Planners use models for many purposes:

1. The first purpose of modeling is simply descriptive: Models are used to depict or represent a complex system more understandably.

2. The second purpose of modeling is analytic, i.e to use rational means to:

   a. Understand and reveal new systems facts and relationships, and to

   b. Use description and analysis of past events experienced in the system to understand and deal with the present, and to project or forecast future ends. In educational planning this means using analysis of past events to assist in the design and implemention of plans, policies, strategies, and programs to meet needs and resolve problems.

   c. To use analysis of the past and present.

3. The third purpose of modeling is to support improved systems decision-making:

   a. The descriptive and analytic advantages of computer-based models allow planners to experiment and learn by testing the effects of simulated changes in the system without having to change real elements at high cost.

   b. The use of the model structure of variables, equations, and coefficients allows planners to probe and test future, possible consequences by forecasting and goal seeking.

4. A fourth pupose of modeling is to improve communication. Models allow planners to discover new relationships in the system and state them in a way that can be readily communicated to decision makers or other members of an education organization who are not technical specialists, but managers or political leaders. To achieve this purpose, models should be (a) easy to learn and use and (b) economical to design and run. The characteristics of a good model are clarity, simplicity, economy, and utility.

An educational planning model describes an element of reality in terms of variables and their values. A constant keeps the same value at all times. A variable varies, which means the values of the variable change over time. A constant will have the same value in the future as it has now, but a variable will change, or take on different values.

Planning deals with the future and the future is change. Thus, variables are of interest in planning for a changing future. We deal every day with the major variables in an education system such as

- Students
- Teachers
- Classrooms
- Books and materials, etc.

The number of teachers may be one variable; the average salary of teachers is a different variable. Enrollment (the number of students) is another variable. A school's enrollment might have been 1000 last year, decreased to 900 this year, and may grow to 1100 next year.

Enrollment, then, is not only a variable--it is also a key variable in schools. Planning identifies key (main) variables and shows how they relate to each other, by using a model.

From the relationship of two key variables to each other, a new third variable may be built. For example, the pupil/teacher ratio is built by dividing enrollments (students) by teachers.

The next section of this chapter will discuss a certain kind of model: Systems picture models.

2.1 THE SIMPLEST MODELS: SYSTEMS PICTURE MODELS

One kind of educational planning model shows how variables are related in order to make a picture of a single school or a system of schools. Picture models show the structure or parts of an institution or school system. This kind of model may show how (1) the faculties and departments and the central administration relate to each other, or how (2) the national Ministry relates to the education offices of the governorates or provinces and how these in turn relate to the school districts and individual schools. The examples on the next three pages show very familiar types of picture models.

1. The Organization Chart

Figure 2.1 (page 4) shows an example of an Organization Chart. The model shows how the offices and officials in the education system in Pakistan are related. Within the Ministry of Education (MOE) it shows line/staff relationships among the secretary, joint secretary, joint education advisor, and other positions in the Ministry.

2. The Systems Chart: Structure

Figure 2.2 (page 5) is a Systems Chart that shows how the various levels and kinds of schools are linked in the Egyptian national system. The chart makes it easy to see the sequence from the nine grades of Primary and Preparatory in Basic Education to the general and special programs in Secondary Education.

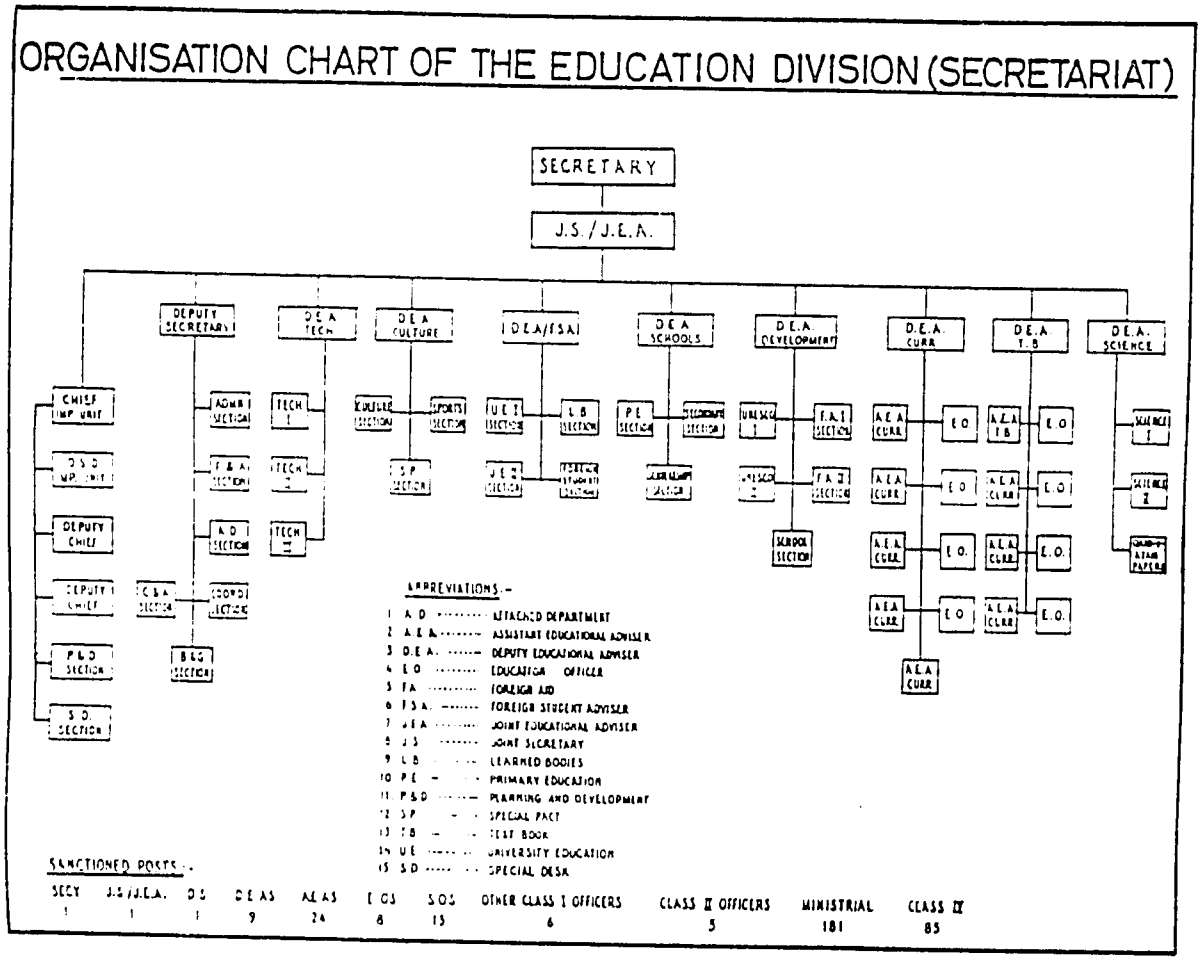
3. The Systems Chart: Process

A model may also show process, or how the parts of the system work. An example of this kind of model would be one showing how annual budget estimates are prepared and sent to the Ministry of Education or Ministry of Finance. Figure 2.3 (page 6) shows a picture model for analyzing the reporting structure for primary/basic education in Egypt.

[Note: The Appendix to this chapter includes a variety of other picture models related to education.]

Figure 2.1

Example of an Organization Chart



Source: Government of Pakistan

Figure 2.2

Example of a Systems Chart Showing Structure

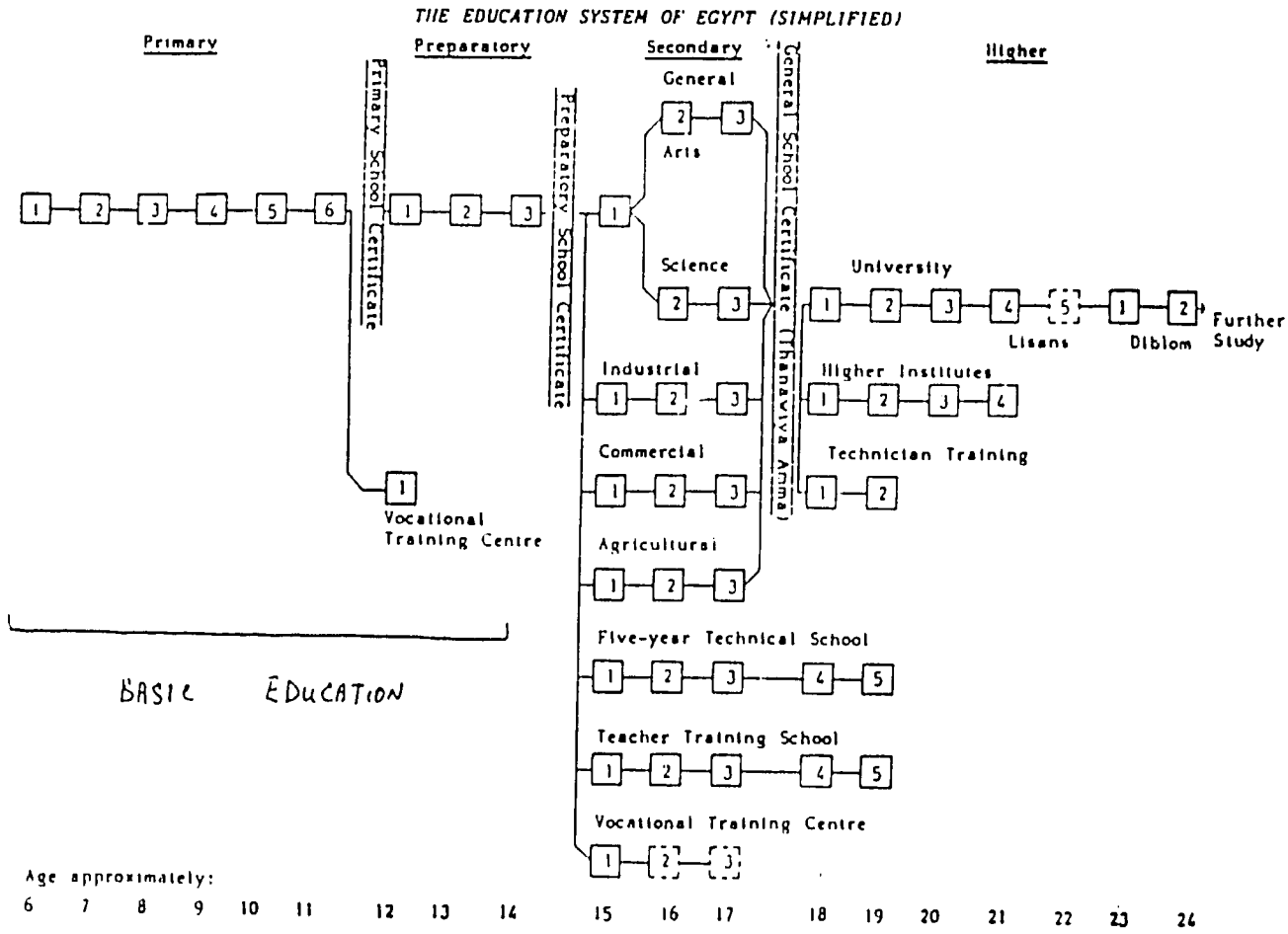
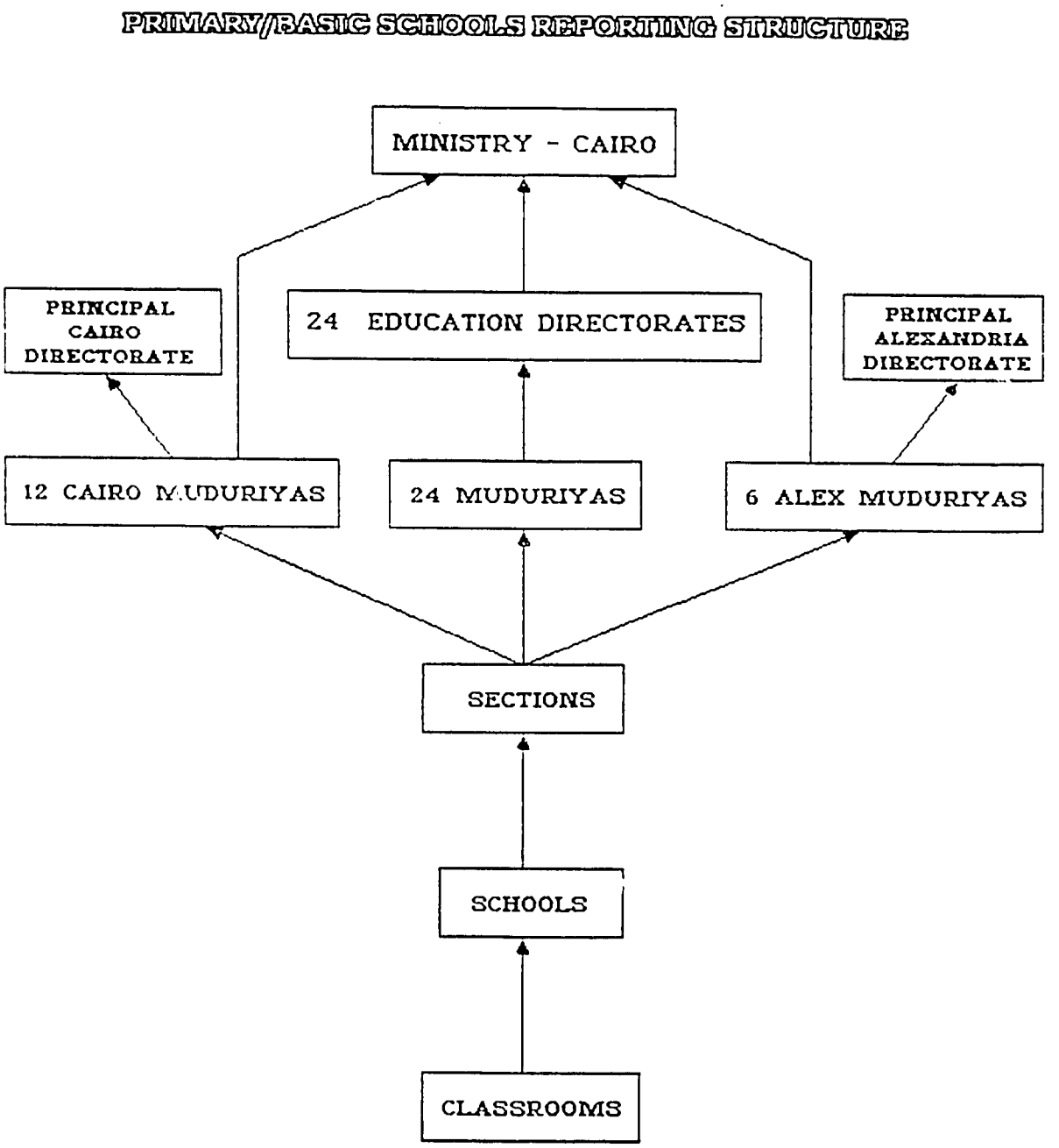


Figure 2.3

Example of a Systems Chart Showing Process



These simple picture models are helpful for understanding the structures and functions in education systems. (Of course, the world of reality is not always the same as the picture models!)

The remainder of this text, however, will deal with algorithmic models. Algorithmic models can help to analyze variables and forecast their future values. These models are different from picture models in three important ways:

1. They are dynamic (They show action overtime),

2. They are quantitative (They use numbers instead of pictures),

3. They lead to solutions.

2.2 ALGORITHMIC MODELS FOR ANALYSIS AND FORECASTING

Algorithmic models often describe relationships with algebraic equations rather than by pictures, arrows, and boxes. For example, here is a basic population forecasting model:

$$P(fut) = P(present) + B - D +/- M$$

This states that the future population [P(fut)] can be forecast from the present population [P(present)] by adding Births, subtracting Deaths, and adding or subtracting Migration (depending on net).

Here is a somewhat less obvious model for forecasting enrollment:

$$E_{(t+1)} = A_{(t)} * E_{(t)} + N_{(t+1)}$$

This states that enrollment next year $E_{(t+1)}$ is the result of enrollment this year $E_{(t)}$ premultiplied by a matrix of flow rates $A_{(t)}$ (promotion, repetition and dropout) from grade to grade and year to year, with new enrollments $N_{(t+1)}$ added in.

Once the goals and objectives are clear, specific targets (descriptions of gaps between actual learning in the schools and the needs and goals of the system) can be specified for accomplishment. For example, suppose a goal for a system states that "80% of the age eligible children will enter the first level of the system". Research on admissions then finds that only 68% of age eligible children are being admitted. Obviously, there is a gap between what is specified in the goal and reality. Planners may at this point set a target such as:

> "Over the next five years, the intake rate of age eligible students entering the first level of the system will be raised from 68% to 80%."

Now the task becomes determining how to meet the target, thereby accomplishing the goals. It is important to realize that the planner does not yet begin to state program objectives. Rather, as shown on Figure 1.1, there are several intermediate stages. These stages are particularly important because goals are often very general statements (and unfortunately sometimes vague) and objectives are specific (and sometimes overly precise). The planner must bridge the gap between the two. This is accomplished with strategies and policies. Strategies and policies are discussed in the next section.

1.4 LINKING GOALS TO OBJECTIVES WITH STRATEGIES AND POLICIES

Strategies lay out in broad terms the alternative ways, the alternative means for attaining goals. Strategies specify how means conflict and interfere, or how means support and help each other. Since all means cannot be pursued at one and the same time, strategies specify the best sequences, the priorities for choosing, deciding on, and implementing the means.

Policies specify (1) how strategies and programs are to be carried out, (2) who will participate in the programs and how, and (3) what the programs will provide and under what conditions. Policies may be written into law or legislation, or in institutional or program regulations.

Rather than discussing abstract definitions, another example will be presented. The example will trace the path between the goals stated in a national plan and the actual implementation of instruction, noting the strategies and policies that make the implementation possible.

The set of goals, policies, and objectives that follows comes from the Second Development Plan of Saudi Arabia (1975-1980). The plan starts with the fundamental goals of society:

1. Goal Statement

Maintain the religious and moral values of Islam
Insure national defense and security
Insure economic development through high growth
Reduce dependence on oil exports and preserve reserves
Foster social stability along with social change
Develop human resources through education/training and health.

2. Strategies, Objectives, Policies

Within the broad strategy of Human Resource Development, objectives, policies, programs, and projects would be developed to cover four major areas:

- o Manpower development
- o Training (Public administration and vocational training)
- o Labor law, services, security
- o Education

Then, each of these four areas must set general objectives for all of its' units. For example, Education, would set general objectives for its' four main branches:

- o Boys' Education
- o Girls' Education
- o Higher Education
- o Religious Education

Within each of these branches, policies and program objectives are set. As an example, Boys' Education will be described in more detail on the next page.

Exercise

SPREADSHEET SOFTWARE FOR MODELING

2.3 EXERCISE OBJECTIVES

The objectives of this exercise are:

1. To examine some of the more common planning models that are built using spreadsheet packages. (LOTUS 1-2-3 will be used for demonstration.) The three main planning tasks to which these models are applied are: (a) Population forecasting, (b) Enrollment forecasting, and (c) Financial forecasting.

2. To review the basic LOTUS spreadsheet skills needed for design and development of the models listed above. This will include:

   (a) How to write labels, values, and formulas on the spreadsheet,

   (b) How to use the basic command menus, and

   (c) How to use a few basic functions like Sum and Copy, using absolute and relative cell addresses.

3. To demonstrate a variety of educational models that have been developed with spreadsheet software.

2.4 BUILDING A SPREADSHEET MODEL: GETTING STARTED

Figure 2.4 below shows the Main Menu choices and the blank spreadsheet structure of LOTUS. (The columns are shown by letters; rows are shown by numbers.) The underlying structure of LOTUS and most spreadsheet modeling systems is the matrix of rows and columns. (Note: This is also the same structure that underlies relational databases like R:Base, but the matrix form is more obvious in the spreadsheet package.)

To get the basic LOTUS menu and the blank spreadsheet on which the model is built:

1. Switch on the computer.
2. Get the systems prompt C > and type: CD\Lotus. (This assumes the LOTUS program is in a sub-directory called Lotus).
3. Next type: 123. This brings the LOTUS logo on screen, followed by the blank screen of the spreadsheet shown in Figure 1. The user can then either begin to write the model on the spreadsheet, or press the / key to get the main menu shown at the top of Figure 2.4.

Figure 2.4

LOTUS Main Menu and Spreadsheet Structure

```
A1:                                                                    MENU
Worksheet  Range  Copy  Move  File  Print  Graph  Data  System  Quit
Global, Insert, Delete, Column, Erase, Titles, Window, Status, Page
        A        B        C        D        E        F        G        H
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
```

Once the blank spreadsheet structure is on the screen, the model will simply be typed into the appropriate cells. Figure 2.5 on the next page shows a simple enrollment and financial planning model developed with LOTUS. It will serve as an illustration for a discussion of the steps of building a model (beginning on page 13).

Figure 2.5

Enrollment and Financial Planning Spreadsheet

(variables) (time periods)

	A	B	C	D	E	F
1	Variables	Period 1	Period 2	Period 3	Period 4	Period 5
2	Grade One	100000				
3	Grade Two	97000				
4	Grade Three	96000				
5	Grade Four	88000				
6	Grade Five	82000				
7	Grade Six	81000				
8	TotEnrol	@Sum(B2..B7)				
		(544,000)				
9	Num Faculty	13614				
10	P/T Ratio	40/1				
11	AvFacSal	5100				
12	ToFacCo	+B9*B11				
		(69,431,400)				
13	Supervision	6843500				
14	InstrMater	1360000				
15	MaintRent	5440000				
16	OtherCosts	2109800				
17	TotExpend	@Sum(B12..B16)				
		(85,184,700)				
18	GovAllocPS	5350				
19	ToGovAlInc	+B9*B18				
		(72,834,900)				
20	Fee Inc	6544350				
21	OthInc	3560000				
22	ToInc	@Sum(B19..B21)				
		(82,939,250)				
23	GR 1GRate	.10				
24	GR2GR	-.11				
25	GovAlPSGr	.08				
26	GRaFacSal	.09				
27	GGraOthCO	.10				
28	FundBal	+B21 - B17				
		(-2,245,450)				

2.5 WRITING THE MODEL ON THE SPREADSHEET

To create the model shown in Figure 2.5, the following steps were followed:

1. The first column was created by writing the label "Variables". In the rows below, the names of the variables in the planning model were typed (from "Grade One" to "FundBal").

2. Column labels were written along the top row ("Period 1" to "Period 5").

3. Opposite the variable "Grade One" and below the column "Period 1", in cell B2 the value of the first year enrollment for Grade One in Period 1 was written: 100000 (Note: Do not use commas, only decimals.)

4. In cell B8, a formula was written which uses the sum function--<u>@SUM(cells summed)</u> in LOTUS. When this formula is entered, it will give the value 544000 (shown in parentheses below cell B8).

5. In Cell B12, a formula was written which instructs the program to multiply the value in Cell B9 by the value in B11. (Note: To tell LOTUS that it is a formula in which the value of B9 is multiplied by B11, a "+" sign is written before B9. Otherwise, the cell entry would be treated as a label.)

6. Other values and labels and formulas are written in other cells. Because Figure 2 shows a small model for teaching purpose, it has 28 variables and it forecasts or projects the values of these variables over five time periods. Obviously, a fully developed model might have many more variables.

7. The final variable was written on the bottom line: "FundBal". "FundBal" is the balance of Total Expenditure and Total Income. In Period 1, the balance shown in parenthesis is negative. The planner can then use the model to forecast enrollments, expenditures, and income over the five year period in order to see what steps are required to attain a positive fund balance--for example, increase the Government allocation per student (called GOVAllocPS on the model).

[Note: Variables are often written in a short form to save space (although worksheet columns can be made wider, using choices Format and Width on the Worksheet menu). Some abbreviations are obvious such as "Supervision" standing for "supervision of instruction costs". But other abbreviations are not as obvious. For example, "on Figure 2, GovAllocPS" stands for "government allocation per student".

Variable names that are too short can be confusing and the model builder may forget what the abbreviated variable name means. Furthermore, abbreviations can impede communication with others who are trying to understand the model. Thus, when building large models, a good practice is to print a list of shortened variable names and their equivalent full name meaning.]

7. The last group of variables shown on Figure 2 are growth rates--in this case arithmetical or percent growth rates per year. Growth rates are multiplied by starting year values to project changes in the basic variables over the periods in the model.

2.6 ANALYZING THE STRUCTURE OF THE MODEL

Before using the model shown in Figure 2.5 to forecast Enrollment, Expenditure, and Income over the time periods, it is worth looking at the figure cell by cell to understand what the model is set up to do in Period 1. The model is a collection of variables that affect Expenditures and Income in a certain school system. Note the following points as you study Figure 2.5:

1. Enrollments in this system determine costs over time.

The ratio of Pupils to Teachers (P/T ratio) is divided into total enrollments to yield the number of teachers required. For example, if the ratio is 30/1 and this is divided into a total enrollment of 6000, then 200 teachers are required. When this number is multiplied by Average Teacher Salary, shown in the model as "AvFacSal", it gives the ToFacCo, or total teacher costs. Other expenditure items for Supervision and Instructional Materials ("InstrMater") or classroom space can be derived as ratios to teachers or basic enrollments. Enrollment is a key variable in the model because it determines (or drives, as planners say) costs.

2. Enrollments in this system also drive income, since the Government allocation is on a per pupil basis.

3. The key variables are linked together in various ways by formulas. As noted above, teaching costs derive from the number of teachers multiplied by average salary costs. However, the number of teachers is related to Enrollment by the Pupil/Teacher ratio. In more complicated models the variables can be related and linked, just as they are in reality. Key variables are linked by using formulas which may be written with any of the 60 or more functions of LOTUS from simple SUM or AVG (Average) to financial functions like FV (Future Value) which raises an amount over a period of years according to a compound rate of interest.

4. Over time, the values in the later periods will be linked to the values in the earlier periods by means of a growth formula. These growth rates can be calculated from past data in the system using LOTUS functions or other methods. Growth rates can then be written into formulas to show how variable values increase or decrease over time.

Having created the basic structure of the model, the next section will discuss using the model for forecasting enrollme

2.7 FORECASTING WITH THE SPREADSHEET MODEL

The value of the spreadsheet for planners is that it can be used to forecast over the time periods with a very simple operation which replicates or copies the formula for a growth rate multiplied by an initial variable value. For example, Figure 2.6 below shows how a formula was written in order to project the Grade One enrollment in Period 1 to Period 2. (The cell value of B2 was written in thousands to save space.) The formula uses the growth rate for Grade One, which was given as .10 in cell B23 of the model (Figure 2). Once this initial formula is entered, the software provides a simple process for replicating the procedure over all periods of the forecast.

Figure 2.6

Forecasting Grade 1 Enrollment Over Time

	A	B	C	D	E	F
1		Period1	Period2	Period3	Period4	Period5
2	Grade One	100	+ B2*(1+$B$23)			
3	etc					

To forecast Grade One enrollment over the five periods, the steps are:

1. Type the formula "+ B2*(1 +$B$23)". A value 110 will appear in C2. (The formula will be explained in detail on the next page.)

2. Press / and select the command Copy.

3. Indicate the Range of Cells to Copy From (in this case C2) where the formula is written and the value 110 is calculated.

4. Indicate the Range of Cells to Copy To (in this case D2...F2).

5. Press Enter and the values are projected from D2 to F2--that is, from period 3 through period 5.

Obviously, the key element in the procedure just described is the formula: + B2*(1 +$B$23). Note the following points about the formula:

A. There is a "+" sign before B2, which means that it is a formula (and its value will be copied over the time periods).

B. There is a "$" sign before the cell value B and 23: "$B$23", but there no "$" before cell entry "B2". This means that:

- o The formula uses the relative value of B2 (It will change over the period of the forecast).
- o The formula uses the absolute value of B23 (It will not change in value over the forecast).

The distinction between absolute cell values and relative cell values is critical in spreadsheet software. In the formula, the enrollment values to be forecast (B2) are relative. This means the program locates a prior value (specified by the location of the cell, rather than a specific value in the cell) and multiplies it by the growth rate. Thus, in each period, the relative cell value will be different:

--In Period 1 it is 100
--In Period 2 it is 100 X 1.1 = 110
--In Period 3 it is 110 X 1.1 = 121
--In Period 4 it is 121 X 1.1 = 133.

By contrast, the same growth rate is used throughout the forecast period, so it must have an absolute cell symbol ($) in front of the cell designation. Thus, for each period, $B$23 will be the same value of 1.1 (1 + .1) because the value is unchanging over time.

The ability to specify absolute and relative values is the powerful feature of LOTUS and other spreadsheets that makes them usable for modeling and forecasting in educational planning. For planners this is perhaps the single most important feature, and it must be understood and mastered to build spreadsheet models.

To summarize this description of the key formula, then, +B2*(1 + $B$23) means that over the range of time periods selected in the Copy command, the value will be the absolute value of the growth rate shown in cell B23 (.1 + 1 =1.1), multiplied by the relative value of C2, then D2, then E2, then F2.

2.8 SUMMARY OF STEPS FOR CREATING THE FORECASTING MODEL

1. Call "LOTUS" or "1-2-3, whichever brings up the program.

2. When the logo appears, press Enter.

3. When the blank worksheet appears with the Ready sign in the top right hand corner, write in the variables, values and formulas in the first period of the model.

4. To forecast the values of any variable from the first period, write a growth formula as shown above, e.g. "+B2*(1+$B$23)". Press Enter to enter the value in cell C2.

5. To project values to future periods, press "/" to get the main menu of LOTUS.

6. Select Worksheet from the main menu, and then select Copy. Press Enter.

7. Indicate the cell location of the formula to be copied From (C2 in the example) and press Enter.

8. Indicate the range of cells to which the relative copy (projection) is to be made To. In the example this range is D3...F2. Press Enter.

9. The value of the variable is projected across the range of cells from D2 to F2, which means time periods 3 - 5.

Using the combination of relative and absolute cell values in the formula "+B2 *(1 + $B$23)" and Copy, LOTUS will simulate a forecast over the five time periods and show these results:

1 **Freshmen**	**100**	**110**	**121**	**133**	**146**

These simple steps make up most of the method for using a spreadsheet model for forecasting. The rest is mastering the commands, functions, and conventions of the specific software package. Logical operators such as @IF (conditions) and especially functions within functions chains (and look up conditions) take practice, but are useful in building models.

However, to build a model that serves much of a purpose in reality requires going beyond the basics into the design and use of macros, and the design of customized menus and screen views, i.e the development of templates and full models. The design of customized templates or models requires a "sense of screen design" (understanding how a segment of the spreadsheet that is about 20 rows and 76 characters wide will look). The rest is a matter of designing layout maps, subject indexes, menus, and macros, and clear instructions.

2.9 ADDITIONAL MATERIAL ON SPREADSHEETS

The basic components just described can be used to develop or generate models of systems, which means representing the main variables of the system, relating these variables in equations, linking them and carrying out operations on variables and equations. In addition, LOTUS (like Multiplan, Javelin, or any spreadsheet system) has other capabilities:

1. Macros can be developed in which series of commands that may be frequently used can be put together so the whole series can be called up and done by calling the macro with one simple command using "/" and the call letter for the macro. Macros can also be built to present the user with special purpose menus designed to fit a particular model application or template.

2. In general, templates are analytically designed and structured to perform some task, usually a sub-component task of a larger model which covers a number of linked planning component tasks. Component tasks of population, enrollment, and manpower forecasts are plan task components that are suited to template design. The templates (or component projections) can then be linked in a large spreadsheet model (although there are efficiency limits on this) using Lotus. The components might more efficiently be built into a comprehensive, multi-modular model, using a more efficient programming language such as Turbo Pascal or HOST/PETS.

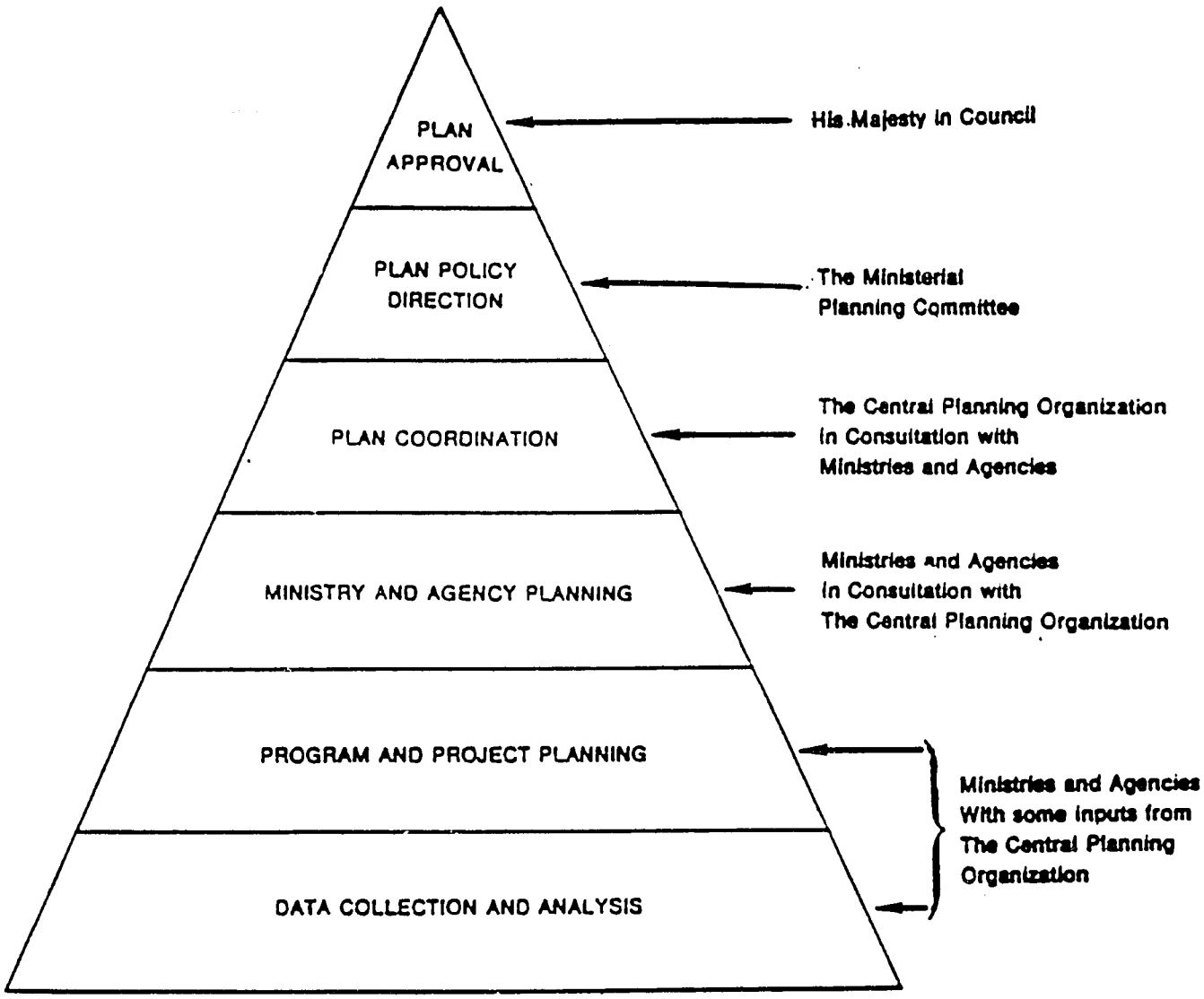
3. Scenarios can be designed with series of what-ifs, and the model recalculated to test the effects of changes. The effect of change in systems variables can be tried by using sensitivity test table arrays. The template components can be linked to form more comprehensive sub-models or models; data can be moved and managed within the model, or imported from other data-bases and sources. Data can also be exported to other models. The spreadsheet can be linked to other modeling and management packages on PC's or mainframe computers.

4. Graphics displays can be designed to enhance descriptive contributions of the model, or to show analytic relationships.

APPENDIX A

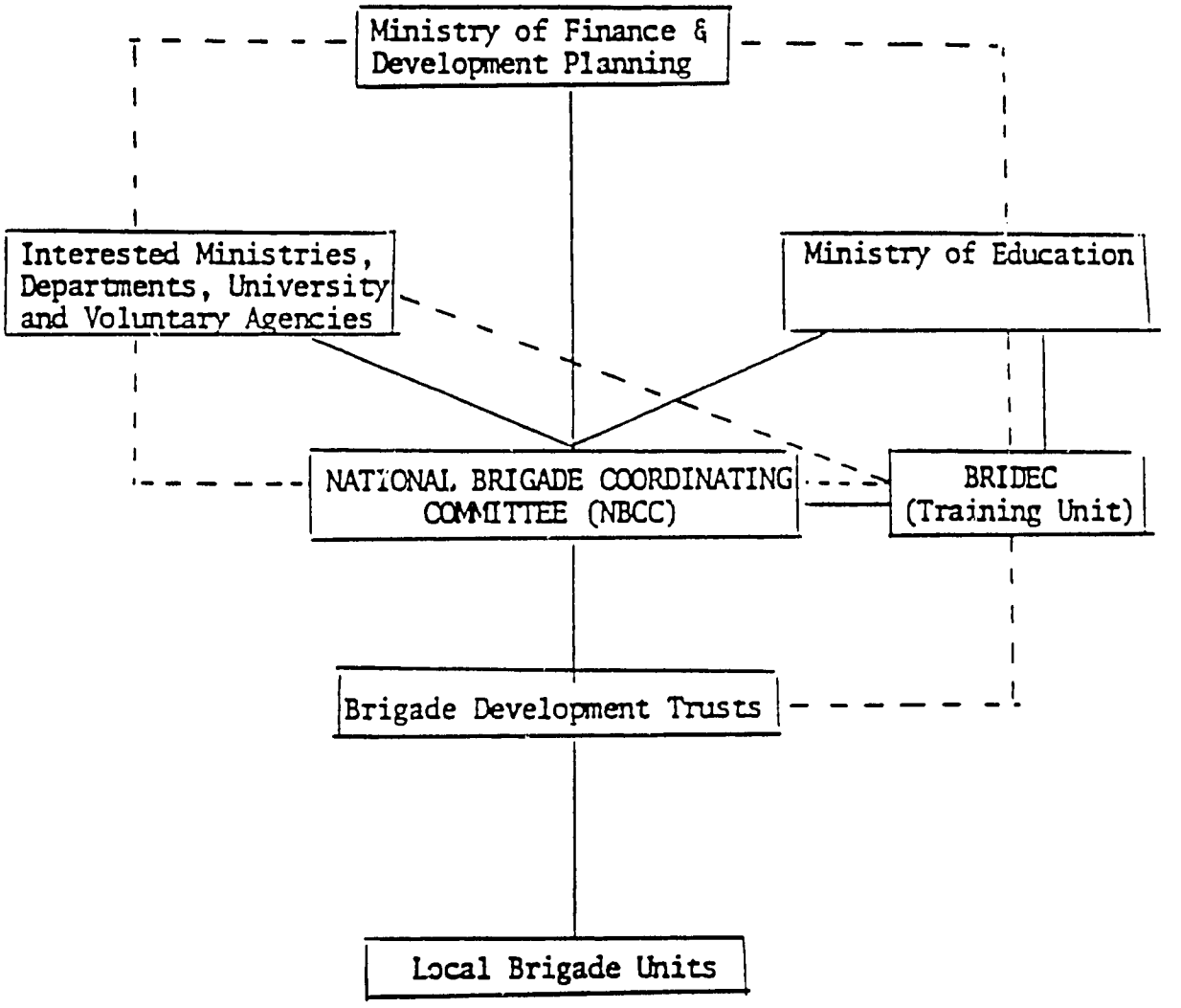
SAMPLES OF MODELS RELATED TO EDUCATION

(Pages 21-26)



THE STRUCTURE OF PLANNING

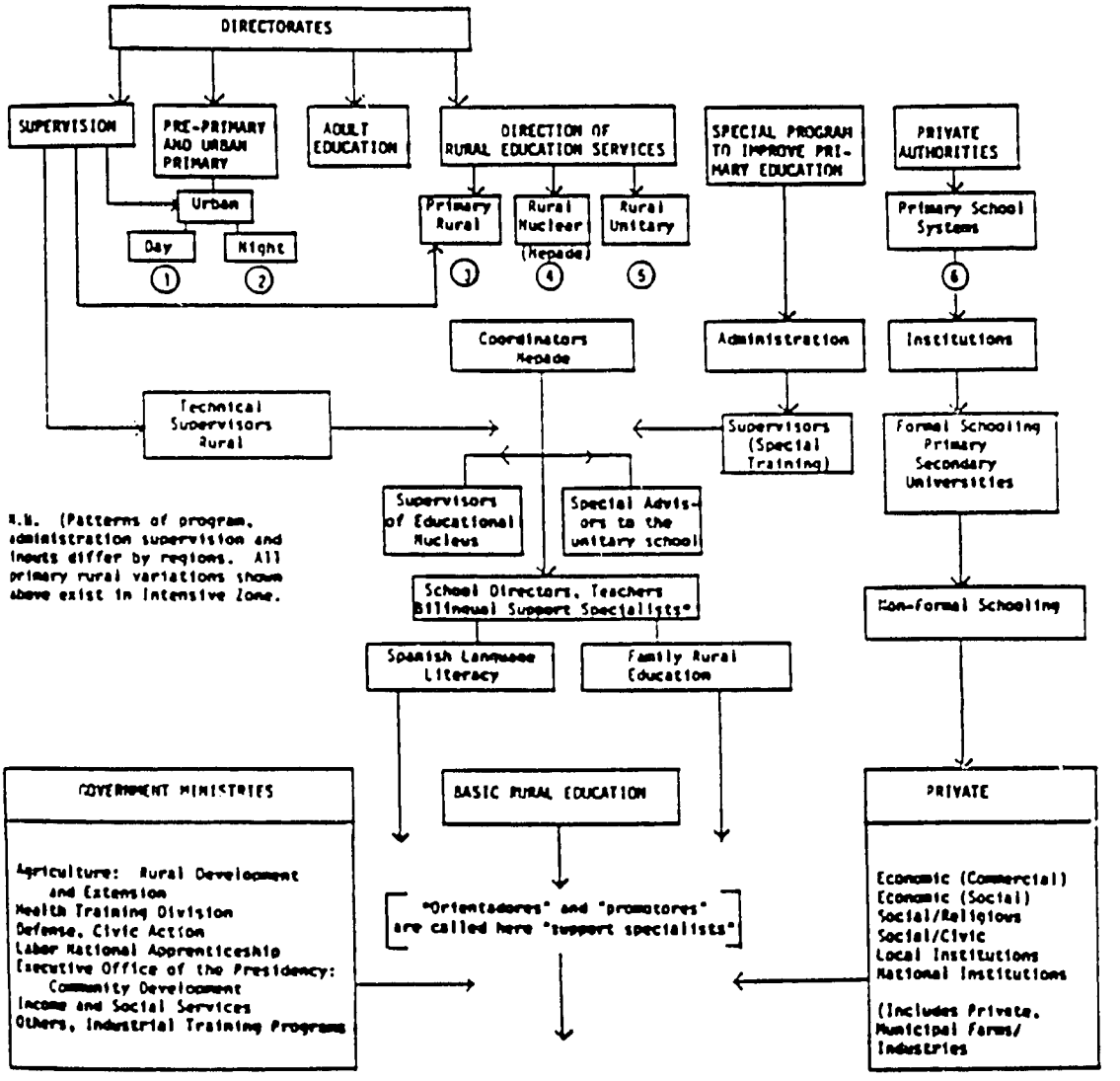
THE BRIGADES: THE NATIONAL ORGANIZATIONAL STRUCTURE



_____ Formal linkages

-------Informal linkages

Actual Patterns of Direction and Supervision of Primary Education, Guatemala

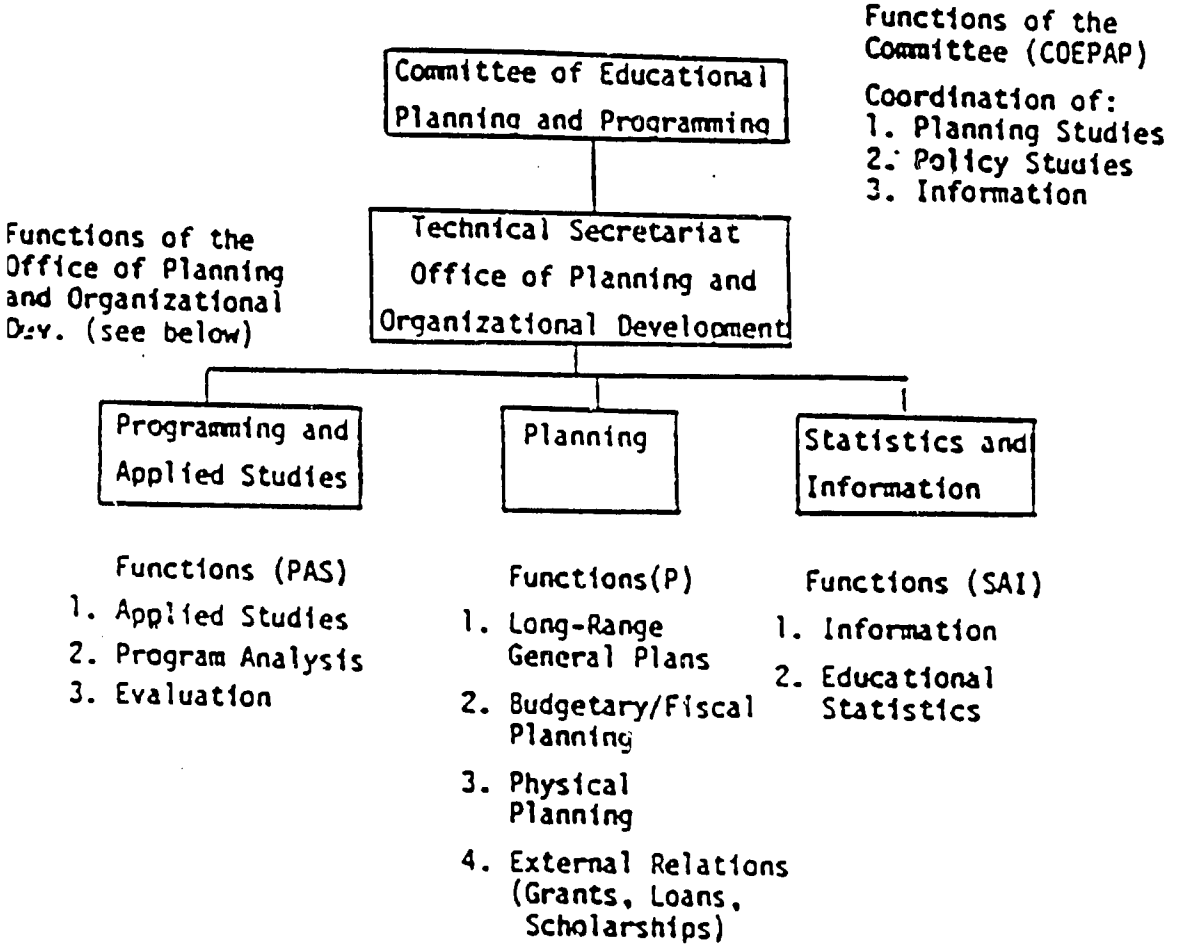


N.B. (Patterns of program, administration supervision and inputs differ by regions. All primary rural variations shown above exist in intensive zone.

Note that there are two urban (1 and 2) and three rural primary school programs (3, 4, and 5). Private schools make a total of six different kinds of rural primary schools, each with different administration, supervision and different programs. There are also nonformal education programs, Literacy and Family Education operating in rural primary schools.

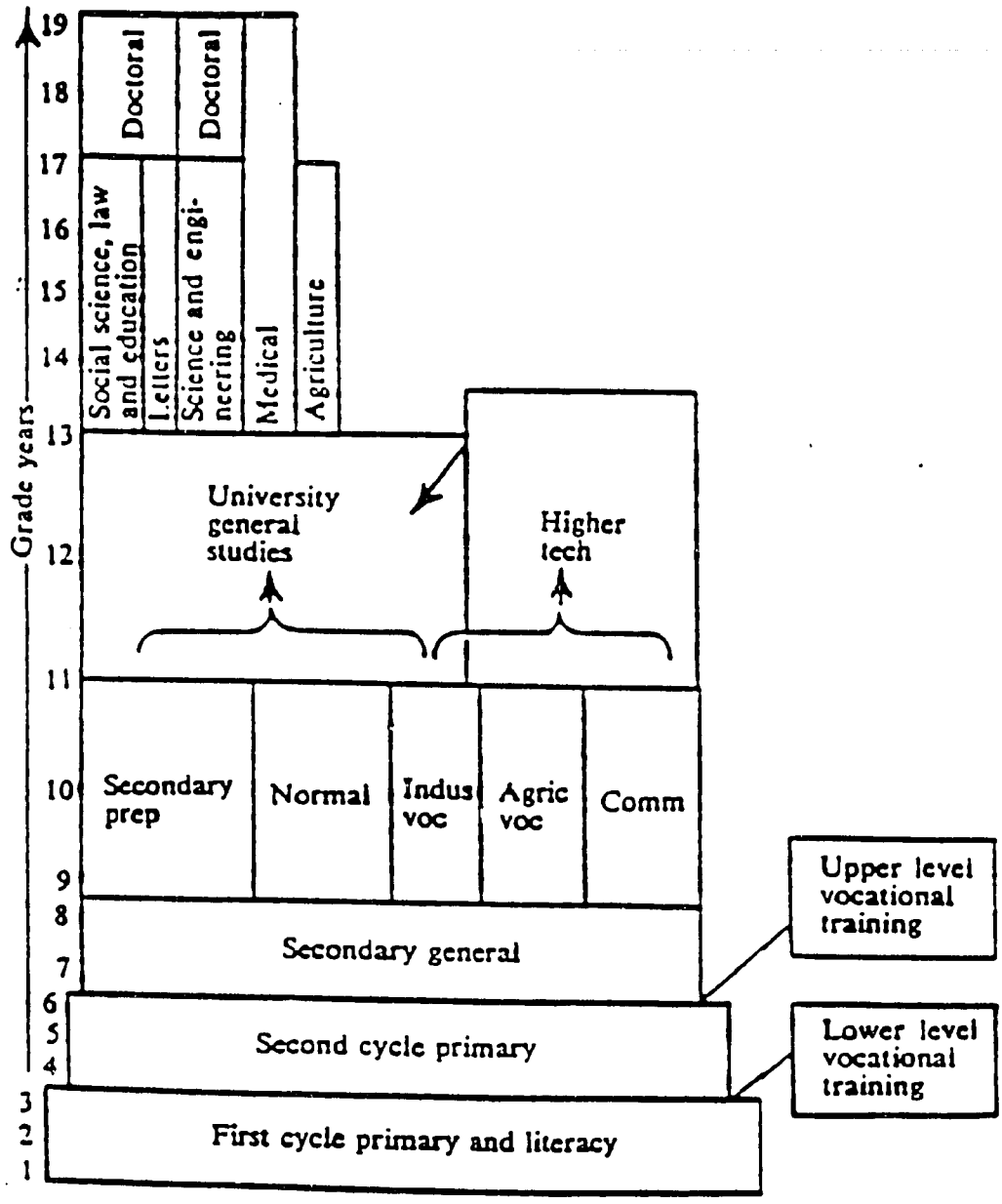
Source: Davis, Russell. "Morphological Analysis Applied to a Systems and Organization Study: The Guatemala Case." USAID/Harvard Project, Paper No. 81, 1979.

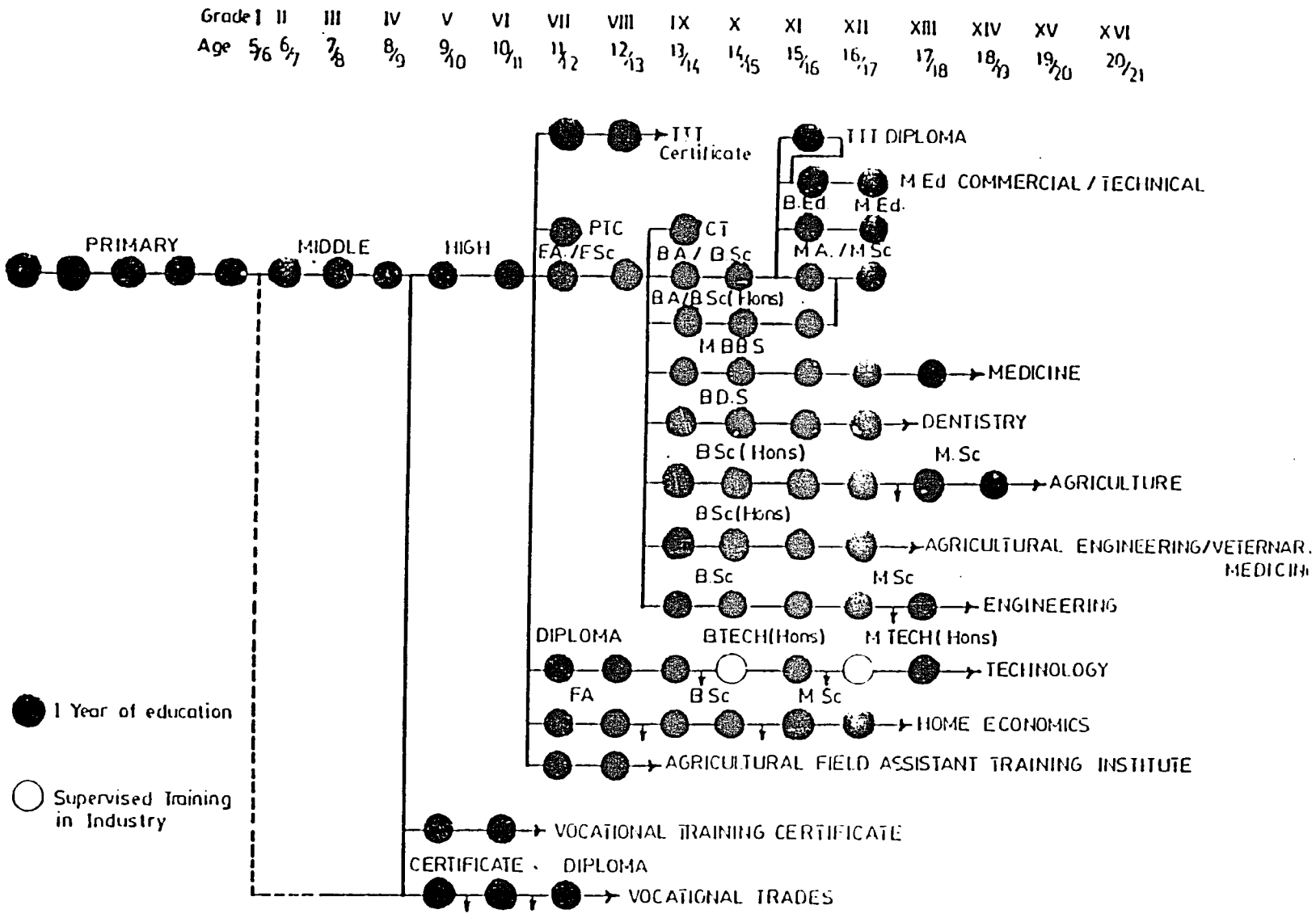
Organizational Schema of the Office (Final Version)



Functions of the Office of Planning and Organizational Development

1. Technical Advice and Information to Committee and President (Minister) that would cover:
   - A. Planning; long and short range, budgetary and physical planning
   - B. Policy Studies
   - C. Program Studies
   - D. Organizational Analysis and Organizational Development
   - E. Information and Public Affairs


Grade years
19
18
17
16
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
Doctoral
Doctoral
Social science, law and education
Letters
Science and engi-neering
Medical
Agriculture
University general studies
Higher tech
Secondary prep
Normal
Indus voc
Agric voc
Comm
Secondary general
Second cycle primary
First cycle primary and literacy
Upper level vocational training
Lower level vocational training


Grade I II III IV V VI VII VIII IX X XI XII XIII XIV XV XVI
Age 5/6 6/7 7/8 8/9 9/10 10/11 11/12 12/13 13/14 14/15 15/16 16/17 17/18 18/19 19/20 20/21
PRIMARY
MIDDLE
HIGH
TTT Certificate
TTT DIPLOMA
M Ed COMMERCIAL / TECHNICAL
B.Ed
M Ed.
PTC
FA/FSc
CT
BA / BSc
MA./MSc
BA/BSc(Hons)
MBBS
MEDICINE
BDS
DENTISTRY
BSc(Hons)
M.Sc
AGRICULTURE
BSc(Hons)
AGRICULTURAL ENGINEERING/VETERNAR. MEDICIN
B.Sc
M Sc
ENGINEERING
DIPLOMA
BTECH(Hons)
M TECH(Hons)
TECHNOLOGY
FA
B Sc
M Sc
HOME ECONOMICS
AGRICULTURAL FIELD ASSISTANT TRAINING INSTITUTE
1 Year of education
Supervised Training in Industry
VOCATIONAL TRAINING CERTIFICATE
CERTIFICATE
DIPLOMA
VOCATIONAL TRADES

APPENDIX B

LIST OF MODELS AND TEMPLATES AVAILABLE FOR DISPLAY

1. The POPEX Model (Cassidy/Davis,1986) forecasts school age population and tests the sensitivity of school entrants forecasts to changes in basic demographic rates. A number of such sensitivity tables and tests can be built up, using the basic template model.

2. The ECC Model (Economics in Curriculum Choice) (Nazareth,1986) is designed to test the effects of changes in costs of school component inputs (teacher, salaries, materials...) on total and unit costs, and to allow substitutions to be made in inputs and to test the effect on costs and financial requirements.

3. EDFISMO (Education Finance Simulation Model (Zymelman, 1987) forecasts the number of students in and flowing through a school system that results from an exogenously derived budget (fraction of government budget allocated to education) and the unit costs of students, which in turn depend on the system characteristics (size, inputs and input norms). It has a student flow model; a model for estimating teacher requirements; a cost sub model; and a budget sub model. Very general rates are assumed and imputed to drive the system over time.

4. NCHEMS templates use macros to structure sub components for financial planning in higher education. The templates cover the main tasks of planning: school population, enrollment, and financial forecasting.

5. EDMODELS (Truesdell, Hyde and Davis, 1978) ARE designed to forecast enrollment flows given flow parameters and entrants. Compiled for use on CMS IBM VM /370; also on VAX at HGSE. Can be used with the population forecasting model DEMOS which is on the same system.

6. HOST/STEP (RTI) is a comprehensive modeling system for enrollment forecasting based on population forecasts and for financial forecasting based on enrollment.

7. GPOPEX6.WK1 A partial LOTUS-based population forecasting model showing only the mortality component for only a few of the population age groups. It illustrates a basic formula which shows survival rates multiplied by the population in an age group in year t who survive to the next age group in the next time period. Figure 1 shows the general outline of the partial model, but to understand it clearly the detail on GPOPEX.WK1 should be examined.

8. HPOCLEX7.WK1 This is a more complete version of the same population forecasting model (#7 above). It shows the demographic component and the fertility component and covers two forecast periods. It is still incomplete. Because it covers several screens of material even in incomplete form it can not be shown in a simple page diagram as Figure 1 was. It must be examined on the Lotus spreadsheet.

9. IPOPMACR.WK1 This is a complete population components forecasting model built with macros on a LOTUS template. It forecasts population and shows the number of students entering into the first grade of a school system. A Lotus sensitivity table has been built to show the extent to which the entrant number are affected by the demographic forecasts.

10. LEGENA9 This model of diagonal flow analysis (with assumptions about repetition and dropout rates) is based on Egyptian primary school enrollments.

11. GrowthFC.WK1 This is a simple model with systems variable(s) projected over time periods according to a growth rate set on the basis or past data and analysis. It is a classic form of the enrollment forecasting component of finanacial planning model (e.g.FPLMOD).

12. Exponent.WK1 Shows the use of LOTUS function Exponent used in calculating growth rate between two time periods to get the growth rates used in the previous GROWTHFC model. Regression can also be used or Logs to fit trend.

13. FloMtrx.WK1 This is the matrix of flow rates pre-multiplied by a column vector of enrollments with new entrants added in.

14. FPLXER.WK1 (also shown on Figure 1) and SMALLMOD.WK1 are partial financial planning spreadsheet models set up to illustrate forecasting Copy. Note: The form of these is similar to GROWTHFC.

15. FiPlMod.WK1 This is a complete financial planning model, used for forecasting using only 28 variables.

16. MACRO.WK1. This illustrates the building of a menu macro.

17. XSCHDB and XXLFACDB are LOTUS-based databases.

Chapter 3

POPULATION FORECASTING

3.0 INTRODUCTION

Educational Planning deals with variables. "Variable" means that the value varies or changes over a time period. The time period that concerns planning is the future. Planning must trace future changes in key educational variables (such as enrollments, number of teachers, or classes) by forecasting. Forecasts show how the values of variables might change in the future. Educational planners forecast

- How external variables (those outside the school system) will change and thus affect education in future years.

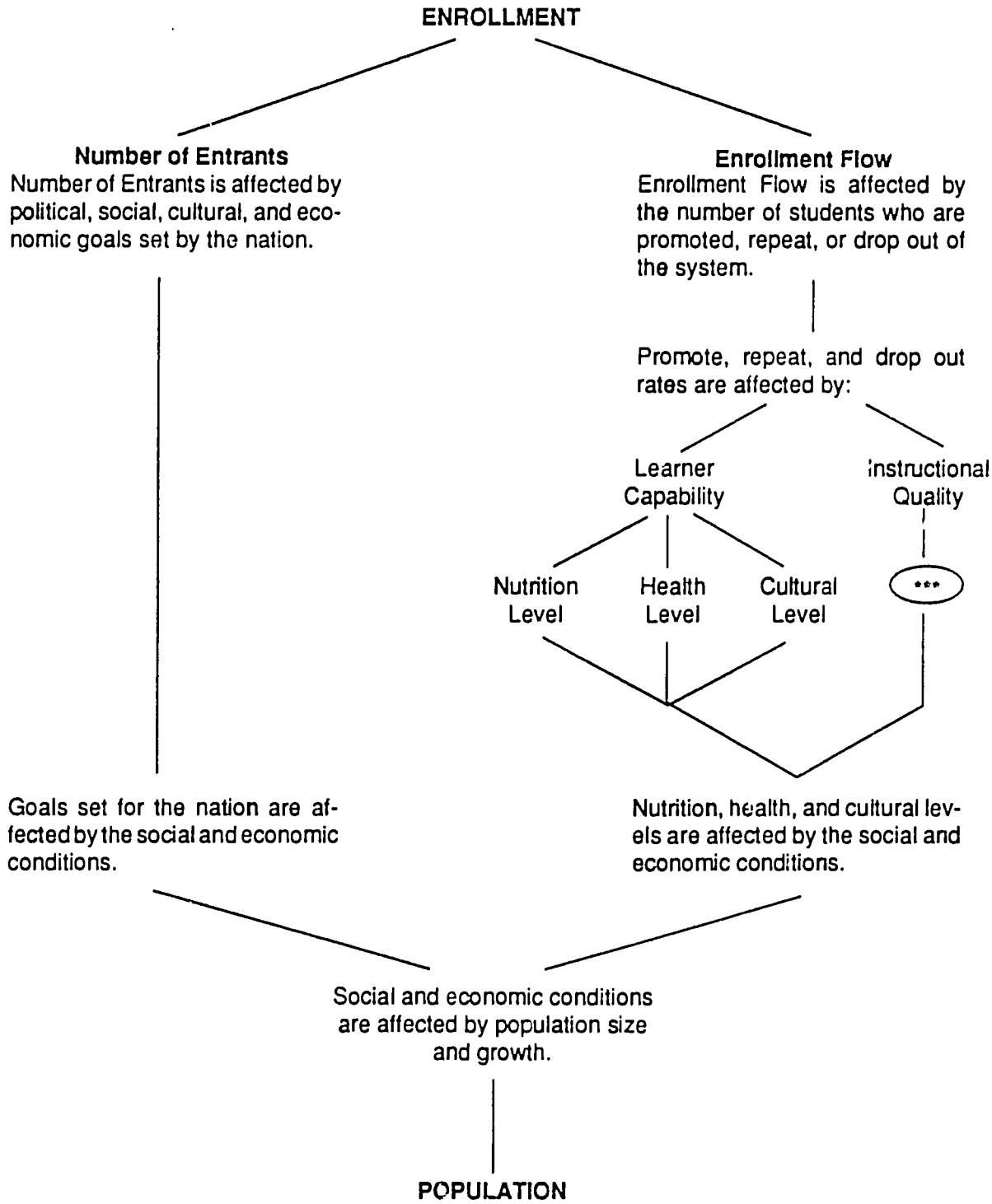
- How internal variables (those inside the school system) will change.

Planners use forecasts of two main kinds of external variables that affect school systems: population growth and economic growth. In terms of population, educational planners are particularly interested in the school-age population, or the number of school-age people classified by age and sex. (Although economic growth is a key variable, the following material will cover only population forecasting.)

The main internal variable of interest to planners is enrollment. Enrollment refers to (1) the number of boys and girls enrolled by grade and (2) the flow of enrollment through the grades of the schools. Chapter 4 will discuss enrollment forecasting.

Planners must deal with both population and enrollment because they are closely related variables. Figure 3.1 on the next page is a simple illustration of this.

Figure 3.1: Relationship Between Enrollment and Population



[***Variables affecting instructional quality are not listed on the chart. What would some of them be?]

Figure 3.1 shows that chains of effects which impact school systems and enrollment can be traced back to the population. Population has a basic effect on social and economic conditions. These conditions then affect education. The quality of education in a nation then affects its' future social and economic conditions, which in turn affects education, and so on.

Thus, it is logical to begin a study of the models and methods of forecasting with a focus on population forecasting. Population forecasting moves planners into the field of demography. Demography is the study of population. Demographic analysis is the study of the numbers and characteristics of populations.

Two characteristics of importance to planners who do forecasts are the age and sex (male/female) of the population. (Another name for sex is gender.) The two examples which follow show very basic population tables:

Basic Population Tables

Table 3.1

Age Group (Year of Age)	Male	Female	Total
0-1	50,000	50,000	100,000
1-2	47,560	47,768	95,328
etc.			

Note that the population is grouped in one year age categories in Table 3.1. Populations may also be grouped in five year age intervals, as in Table 3.2 on the next page.

Table 3.2

Age Group	Male	Female	Total
0-4	247,860	243,970	491,830
5-9	232,800	229,640	462,440
10-14	.	.	.
etc.	.	.	.

Population forecasts are the basis of all social and educational planning. The educational goals and targets discussed in the first chapter often include population statistics. This is true of the following goal:

"All children of school age (6-11) will receive a complete primary school education (grades 1-6) that will provide the foundation for future citizenship and productive employment."

Once a goal has been stated and the number of children in school age in the population is known, specific targets for entrance into school are set. The three targets below include very precise population statistics:

"In 1988, 178 six-year-old boys and 187 six-year-old girls will be admitted into grade 1 in the District."

"In the Province (Governorate), 17,876 boys and 18,972 girls will reach the six-year-old admission age."

"In the second year of the Fifth Plan, 1.2 million children will be admitted to the national schools."

If population estimates are so important to educational planning, how are they determined? The following section will introduce the basic components and methods for population forecasting.

3.1 POPULATION FORECASTING: COMPONENTS AND BASIC MODEL

Four components influence the future size and characteristics of a population:

1. Base Year Population

This means the numbers of males and females in each age group (0-4, 5-9, etc.) in the current population.

2. Fertility

The numbers of males and females added to the population from <u>births</u>. This component is called <u>Fertility</u>.

3. Mortality

The numbers and characteristics of those going out of the population from <u>deaths</u>. This is the <u>Mortality</u> component.

4. Migration

The numbers by age and sex of persons who come into the population (in all ages) from another population.

The basic model for population forecasting makes use of these four components. It begins with a base year, or current population. Births are added. Deaths are subtracted. Migration may be added or subtracted. If more people come into the population than go out migration is plus; if more go out it is a minus.

In other words, the future, or the forecast population equals the present population (the base year), plus births, minus deaths, plus or minus migration. This statement is expressed in symbols on the next page.

Basic Model for
Population Forecasting

$$FP = PP + B - D +/- \text{Migration}$$

Where:

Future Population	=	FP	
Present Population	=	PP	(Base year population component)
Births	=	B	(Fertility component)
Deaths	=	D	(Mortality component)
Migration	=	M	(Migration component)

To show how the four components are actually calculated and used in a forecasting model, an example of a national population forecast will be presented.

3.2 NATIONAL POPULATION FORECAST EXAMPLE

This example will forecast a national population over a 10 year period in two five-year steps: 1985 to 1990 and 1990 to 1995. The data for population forecasting usually comes from two main sources:

- Data on the <u>base year population</u> is taken from a current census or from a past census brought forward and up to date.
- Data on <u>births</u>, <u>deaths</u>, and <u>migration</u> is usually taken from public registers or records.

The next four sections will describe how each of the four components are determined for the forecast.

1. The Base Year Population

Table 3.3 shows the base year population for this example by age and sex. Note that it uses five year groupings for age (which is the usual form for population forecasts). The table has been shortened for the example: The age groups from 20 - 74 have been left out.

Each five year age group will be forecast ahead for five years. The 0-4 age group in 1985 will be the 5-9 group in 1990. (Five year groups can be divided into single year groups using a standard set of demographic tables called Sprague Multipliers. See Appendix A at the end of this chapter for more information on Sprague Multipliers.)

Table 3.3

Base Year Population Table (1985)

Age Group	Male	Female	Total
0-4	475,291	459,374	.
5-9	402,120	392,669	
10-14	314,449	340,742	
15-19	302,190	315,684	
.			
.			
.			
75-80	2,382	4,030	
80+	1,093	1,718	

Even this simple table (based on actual data) shows some demographic truths: There are more males in early age groups, but more females live to old age. This is true of populations in all countries, unless some external event (such as war, famine, or disease) acts to change the general pattern.

2. The Fertility Component

In this example, the birth data from past years used to calculate fertility rates will not be shown. However, it is good to know that demographers analyze past trends in variables that affect fertility rates in order to make estimates of how birth and fertility rates will change in future years. For example, rural to urban migration, rising education of females, and increased participation of females in the work force all act to lower birth rates and fertility.

Based on their analysis of past data and trends, demographers make trial estimates (called hypotheses). In forecasts, different or alternative sets of hypotheses may be run. Planners who use the forecasts may choose the alternative hypothesis that seems most likely to them. Forecasts rarely show single sets of results since the future is not certain.

Table 3.4 and Table 3.5 on the next page show two different hypotheses concerning fertility rates for women by specific age groups, from 15-19 through 45-49. (These are considered the fertile age groups.) Demographers call these age-specific fertility rates. Note the similarities in the tables. Both show that fertility rates:

(a) Differ by age,

(b) Start low at the youngest age group,

(c) Rise to a high point for the age group 25-29 when women have most children, and

(d) Decline in later years as fertility declines with age.

The difference between the two tables is that Hypothesis I (Table 3.4) states that rates will stay the same over the forecast years (1990 and 1995). Hypothesis II (Table 3.5) projects that the fertility rates will decline over the forecast years.

Table 3.4

Births per 1000 Women in Age Group
Hypothesis I: Fertility Rates Unchanged Over Forecast

Female Age Group	1980	1985*	1990	1995
15-19	84.4	84.4	same	same
20-24	223.3	223.3		
25-29	260.4	260.4		
30-34	198.7	198.7		
35-39	139.0	139.0		
40-44	64.9	64.9		
45-49	15.8	15.8		

Table 3.5

Births per 1000 Women in Age Group
Hypothesis II: Rates Declining Over Forecast

Female Age Group	1980	1985*	1990	1995
15-19	84.4	81.3	78.4	75.7
20-24	223.3	214.3	206.7	199.4
25-29	260.4	251.3	242.2	233.6
30-34	198.7	191.5	184.9	178.3
35-39	139.0	134.0	129.2	124.6
40-45	64.9	62.6	60.4	58.3
45-49	15.8	15.2	14.7	14.1

*1985 = base year for the forecast

[Note: To convert births per 1000 into a fertility rate in the form of a decimal fraction, divide the figures in Tables 3.4 or 3.5 by 1000. For example, 84.4 / 1000 = a fertility rate of .0844.]

Another factor to consider in relation to fertility is the birth sex ratio. Males and females are not born in equal numbers. In almost all populations of the world, more males are born:

Table 3.6

Birth Ratios for Sexes

Male Birth Ratio	Female Birth Ratio
.5089	.4911

3. The Mortality Component

For this example, the data and analysis on which mortality is estimated will not be shown. However, the method will be generally described.

A. Compute death rates (m)

On the basis of registered deaths for each age/sex group, age specific death rates (m's) are computed for each age group. The rates show the number of deaths for each 1000 (or 100,000) in the age group. Infant death rates (deaths in the 0-1 year old group) are almost always computed and tabled separately.

B. Compute probability of dying (q)

Once age specific death rates (m) are computed, they are converted into the probability of dying within any age group interval. The probability of dying is symbolized as "q".

C. Compute probability of survival (p)

If the probability of dying at a given age is q, then the probability of surviving is: 1 - q. (Why does this always work?) The probability of surviving is symbolized as "p".

D. Estimate mortality rates

After determining the m, q, and p rates, demographers build tables--called Life Tables--which show the probability of dying (qx) or surviving (px) to any given age for males and females. (See Appendix B at the end of this chapter for more information on life tables.) Planners, then, use the death and survival rates from life tables to estimate the mortality rates for any given population.

Demographers also build life tables called Model Life Tables which show death and survival rates for many different populations with widely different death rates. Using age specific death rates and the probability of surviving, planners can choose the levels in the Model Life Table that fit the population for which they are forecasting. This level then determines the mortality component of the forecast.

Table 3.7 on the next page is a Model Life Table prepared by the United Nations. It shows a life table for males and females with 24 levels of survival rates. The table starts with Level 0 with a high mortality and low survival rate and goes up to level 115, with low mortality and high survival rates. (Only levels 60-115 are shown in Table 6.)

For this example, the survival ratios (taken from Table 3.7) will rise one level during each period of the forecast:

	1985-1990	1990-1995	1995-2000
Model Life Table Levels:	80	85	90

Table 3.7

UN Model Life Tables (Survival ratios)

SEX & AGE in YEARS	LEVEL 60	LEVEL 65	LEVEL 70	LEVEL 75	LEVEL 80	LEVEL 85	LEVEL 90	LEVEL 95	LEVEL 100	LEVEL 105	LEVEL 110	LEVEL 115
	(AA 50.0)	(AA 52.5)	(AA 55.0)	(AA 57.6)	(AA 60.4)	(AA 63.2)	(AA 65.8)	(AA 68.2)	(AA 70.2)	(AA 71.7)	(AA 73.0)	(AA 73.9)
MALES												
(BIRTHS)	0.8406	0.8557	0.8703	0.8877	0.9070	0.9262	0.9438	0.9580	0.9678	0.9744	0.9788	0.9817
0-4	0.9445	0.9518	0.9584	0.9648	0.9708	0.9765	0.9818	0.9867	0.9908	0.9963	0.9954	0.9966
5-9	0.9834	0.9856	0.9876	0.9893	0.9909	0.9924	0.9937	0.9949	0.9963	0.9974	0.9982	0.9986
10-14	0.9827	0.9848	0.9867	0.9884	0.9900	0.9914	0.9928	0.9941	0.9955	0.9968	0.9977	0.9982
15-19	0.9741	0.9772	0.9798	0.9824	0.9848	0.9871	0.9893	0.9913	0.9935	0.9953	0.9965	0.9974
20-24	0.9688	0.9725	0.9758	0.9789	0.9819	0.9847	0.9873	0.9898	0.9924	0.9944	0.9957	0.9966
25-29	0.9673	0.9712	0.9748	0.9781	0.9811	0.9840	0.9866	0.9891	0.9916	0.9935	0.9948	0.9956
30-34	0.9644	0.9686	0.9725	0.9760	0.9792	0.9822	0.9849	0.9874	0.9897	0.9917	0.9930	0.9939
35-39	0.9580	0.9628	0.9671	0.9711	0.9746	0.9779	0.9808	0.9835	0.9860	0.9881	0.9897	0.9907
40-44	0.9467	0.9523	0.9575	0.9617	0.9658	0.9696	0.9728	0.9759	0.9786	0.9812	0.9835	0.9851
45-49	0.9294	0.9359	0.9419	0.9469	0.9516	0.9558	0.9596	0.9632	0.9664	0.9696	0.9727	0.9752
50-54	0.9045	0.9121	0.9189	0.9250	0.9306	0.9356	0.9400	0.9441	0.9481	0.9521	0.9558	0.9594
55-59	0.8678	0.8767	0.8847	0.8919	0.8985	0.9044	0.9098	0.9148	0.9198	0.9249	0.9296	0.9341
60-64	0.8141	0.8244	0.8338	0.8423	0.8502	0.8573	0.8639	0.8702	0.8765	0.8830	0.8890	0.8945
65-69	0.7367	0.7487	0.7598	0.7698	0.7794	0.7881	0.7963	0.8042	0.8119	0.8197	0.8269	0.8335
70-74	0.6334	0.6469	0.6594	0.6708	0.6816	0.6917	0.7012	0.7102	0.7192	0.7283	0.7372	0.7454
75-79	0.5049	0.5193	0.5326	0.5449	0.5566	0.5675	0.5779	0.5878	0.5976	0.6072	0.6176	0.6280
80+	0.2988	0.3091	0.3186	0.3272	0.3352	0.3426	0.3493	0.3556	0.3613	0.3666	0.3721	0.3775
FEMALES												
(BIRTHS)	0.8594	0.8739	0.8882	0.9036	0.9208	0.9380	0.9535	0.9660	0.9744	0.9801	0.9838	0.9864
0-4	0.9463	0.9537	0.9607	0.9669	0.9731	0.9791	0.9844	0.9892	0.9925	0.9948	0.9963	0.9973
5-9	0.9828	0.9852	0.9874	0.9895	0.9914	0.9932	0.9948	0.9962	0.9972	0.9981	0.9986	0.9990
10-14	0.9819	0.9843	0.9865	0.9886	0.9906	0.9925	0.9941	0.9956	0.9967	0.9976	0.9982	0.9987
15-19	0.9745	0.9778	0.9809	0.9838	0.9865	0.9891	0.9914	0.9935	0.9952	0.9965	0.9974	0.9980
20-24	0.9594	0.9735	0.9771	0.9807	0.9839	0.9869	0.9896	0.9920	0.9940	0.9956	0.9966	0.9973
25-29	0.9674	0.9718	0.9756	0.9792	0.9826	0.9858	0.9884	0.9909	0.9929	0.9945	0.9956	0.9963
30-34	0.9654	0.9700	0.9739	0.9777	0.9811	0.9842	0.9868	0.9893	0.9913	0.9929	0.9941	0.9948
35-39	0.9622	0.9669	0.9710	0.9749	0.9783	0.9814	0.9841	0.9866	0.9886	0.9903	0.9915	0.9923
40-44	0.9556	0.9606	0.9650	0.9691	0.9727	0.9759	0.9788	0.9814	0.9835	0.9853	0.9868	0.9879
45-49	0.9437	0.9493	0.9543	0.9589	0.9631	0.9668	0.9700	0.9731	0.9754	0.9775	0.9792	0.9806
50-54	0.9252	0.9318	0.9377	0.9433	0.9482	0.9529	0.9570	0.9609	0.9637	0.9661	0.9680	0.9695
55-59	0.8949	0.9030	0.9104	0.9175	0.9238	0.9298	0.9351	0.9402	0.9437	0.9466	0.9491	0.9510
60-64	0.8459	0.8559	0.8653	0.8743	0.8825	0.8903	0.8971	0.9036	0.9083	0.9122	0.9161	0.9196
65-69	0.7710	0.7832	0.7948	0.8059	0.8162	0.8259	0.8345	0.8429	0.8490	0.8549	0.8609	0.8668
70-74	0.6662	0.6804	0.6940	0.7073	0.7197	0.7316	0.7422	0.7525	0.7602	0.7677	0.7764	0.7852
75-79	0.5369	0.5523	0.5672	0.5819	0.5955	0.6088	0.6207	0.6323	0.6410	0.6493	0.6592	0.6697
80+	0.3211	0.3313	0.3406	0.3495	0.3576	0.3652	0.3719	0.3781	0.3828	0.3869	0.3913	0.3957

4. The Migration Component

The final component to consider is the migration component. In national level forecasts, net migration is small and this component may be left out of the forecast. However, in sub-national level forecasts for provinces, states, or local districts, the migration component may take a great deal of effort and analysis, particularly when the internal migration is high and the data poor.

This example is of a national level forecast. Thus, for purposes of the example, the estimate of migration is 0.

Migration Estimate:

Net Migration will be 0 over the 15 year forecast

Now that the four components have been estimated, the next step is to insert them into the basic model for population forecasting. Earlier in this chapter, the formula for the basic population forecasting model was stated this way:

FP = PP + B - D +/- Migration

The formula can now be read more precisely by interpreting the symbols in terms of the four components:

FP = Future Population
PP = Present Population
B = Fertility Component
D = Mortality Component
M = Migration Component

The next section of this chapter will describe how the four components of the example are used in the basic enrollment forecasting model.

3.3 USING THE ENROLLMENT FORECASTING MODEL

Once the components are estimated, the forecast method is simple. The example continues with Table 3.8 on the next page which shows a partial forecast of the female population from the base year 1985 to the years 1990 and 1995. For each successive five year period (called a quinquennium in demography), the numbers in each five year age group are "survived" into the next quinquennium on the basis of the mortality component. All the age groups can be estimated this way except for the 0-4 group. This group results from births and must be handled differently.

Table 3.8 shows figures for only the first two forecast periods. However, forecasting to a third period (to the year 2000) would be done exactly the same way. However, forecasting beyond a 10 year period would be of doubtful worth, since changes in the basic demographic components could not be foreseen. In most cases forecasts are constantly checked, data and assumptions are changed frequently, and the forecast is pushed ahead for another five or ten years only when these precautions have taken place.

The P's on Table 3.8 are survival ratios taken from the appropriate mortality level of the UN Model Life on page 12 (Table 3.7). The P's show the proportion of the age group that survive from one age group to the next age group over the five year period of each cycle of the forecast. For example, the group that is 0-4 in the base year 1985 is multiplied by the survival rate for that age group at Level 80. The result becomes the number that will be 5-9 in 1990:

459,374 X .9731 = 447,017

This number then moves down one row to the next age group in the table. To determine how many of this age group will survive to the next category (10-14 in 1995), 447,017 is multiplied by the 1990 survival P for the 5-9 age group:

447,017 X .9932 = 443,977

The numbers below the diagonal can all be filled in the same way. After studying Table 3.8, fill in several more of the blanks to be sure you understand the process.

Table 3.8

Forecast Method Partially Shown: Females, 1990 and 1995

Age Group	1985 Pop.	Survival P*	1990 Pop.	Survival P**	1995 Pop.
0-4	459,374	0.9731		0.9791	
5-9	392,669	0.9914	447,017	0.9932	
10-14	340,742	0.9906	389,292	0.9925	443,977
15-19	315,684	0.9865	337,539	0.9891	386,372
20-24	311,429	0.9839	311,422	.	.
25-29	289,619	0.9826	306,415	.	.
.					
.					
80+	1,718				

P* Female survival rates from Level 80 of UN Model (Table 3.7)

P** Female survival rates from Level 85 showing a slight improvement.

[Note: Only one mortality component is used in this example. It is usual to do alternative sets of projections, a practice greatly simplified by using computers.]

The numbers below the line on Table 3.8 can be filled in for all ages for all future years of the forecast. However, to calculate the numbers <u>above</u> the line (including the number in the 0-4 age group in 1990), the <u>fertility component</u> must be used to estimate births.

Table 3.9 on the next page shows how new birth rates are calculated. This short example will cover only births 1985-1990 and work only with female children. However, to include males requires only two more columns and the same method. Explanations of the calculations appear below the table.

Table 3.9

Projection of Births and Survivors for Period 1985-1990

Age	(1) Females	(2) 5 Year Fertil. Rate	(3) Total Births	(4) Female Births	(5) Number Surviving	Etc.
15-19	326,612	.4065	132,768	65,202	60,038	
20-24	311.426	1.0715	333,693	163,877	150,897	
25-29	298.017	1.2565	374,458	183,896	169,332	
30-34	223,715	.9575	214,207	105,197	96,865	
35-39	199,583	.6700	133,721	65,670	60,469	
40-44	175,226	.3130	54,846	26,935	24,802	
45-49	145,662	.0760	11,070	5,436	5,006	
					567,409	

(1) Females

This is the average number of women in each age group in 1985 and 1990. (The figures come from Table 3.8.) The figures for 1985 and 1990 are summed and divided by 2 to get the average, e.g. **315684 + 337539 = 653223/2 = 326612.**

(2) Five Year Fertility Rate

The five year fertility rate results from multiplying the one year rate in Table 3.5 by 5 to get the one year rate. The one year rate for 1985 = .0813. Hence: .0813 X 5 = .4065--the five year rate.

(3) Total Births

Multiply (1) by (2).

(4) Female Births

Total births multiplied by female birth fraction (Table 3.6) to get female births: **132768 X .4911 = 65202.**

(5) Females Surviving

Female Births multiplied by survival rate at birth taken from level 80 of UN Life Table (Table 3.7): 65202 X .9208 = 60038 surviving females births born to mothers in the 15-19 age group.

Summing column (5) in Table 3.9 will give the projected 0-4 female population for 1990 to be filled in on Table 3.8. Most of the same procedure must be followed for male births. (Ask yourself, with which column do we begin to compute male?) Computing these figures and adding them to the number of female births will give a final 0-4 population forecast for 1990.

Summary: Importance of Population Forecasting (and a Caution)

Determining the present and future school age population is essential in planning for three reasons:

1. The difference between the population eligible to attend school and the actual school enrollments is a major indicator of how a school is performing. The term coverage refers to the percent of the eligible age group that is actually served by the schools. Coverage, then, shows whether or not school systems are attaining their goals and targets.

2. Population forecasts show the number who reach school entrance age now and in future years. The current number of entrants can be used to set future targets. Attainment of these future targets will serve as another indicator of school system performance.

3. Forecasts of school age populations serve as the basis of educational planning by providing the basic components for enrollment forecasting, the topic of Chapter 4.

The basic data necessary for population forecasting--the population, births, deaths, and migrations rates--are easy to calculate. What is not so easy, however, is estimating how these rates will change in the future. No matter how careful the planning, no one can know the future; it is full of uncertainties. It is important to remember that forecasts are based on judgments about how births,

deaths, and migration will increase or decrease over future years. These judgments should not be mere guesses, of course. They should be based on analysis of how the key demographic rates--mortality and fertility--have changed in the past. They are also based on how other variables in the system that influence these rates have changed in the past. These judgements are built into systematic sets, called assumptions. The key to population forecasts (or any forecasts) is the strength of the assumptions. Experienced planners realize that the accuracy of a forecast depends on the accuracy of the demographic assumptions.

APPENDIX A

INFORMATION ON SPRAGUE MULTIPLIER PANELS

The example in Chapter 3 was a population forecast of five year age groups in five year periods. Educational planners usually require populations in single years in order to estimate the age-eligible pool of entrants, to calculate the percent of school age population enrolled in school, or to get the percent of age groups enrolled in various grades. Demographers usually do forecasts in five year age groups. They then distribute the age groups into single years using Sprague Multiplier Panels. These are tables that apportion five year grouped populations into single years for all age groups.

Table 3.10 on the next page shows Sprague Multiplier Panels. The panels use different proportions to distribute five year totals in the various five year age groups of a population. The first end panel N1 stands for the 0-4 group; N2 for 5-9 etc. The lower case n's are the proportions of the single year age groups.

Table 3.11 shows how the total in age group 15-19 are apportioned into the single years 15, 16, 17, 18, and 19.

Table 3.10

The Sprague Multipliers

	N_1	N_2	N_3	N_4	N_5
			FIRST END-PANEL		
n_1.....	+0.3616	−0.2768	+0.1488	−0.0336	—
n_2.....	+0.2640	−0.0960	+0.0400	−0.0080	—
n_3.....	+0.1840	+0.0400	−0.0320	+0.0080	—
n_4.....	+0.1200	+0.1360	−0.0720	+0.0160	—
n_5.....	+0.0704	+0.1968	−0.0848	+0.0176	—
			FIRST NEXT-TO-END PANEL		
n_1.....	+0.0336	+0.2272	−0.0752	+0.0144	—
n_2.....	+0.0080	+0.2320	−0.0480	+0.0080	—
n_3.....	−0.0080	+0.2160	−0.0080	+0.0000	—
n_4.....	−0.0160	+0.1840	+0.0400	−0.0080	—
n_5.....	−0.0176	+0.1408	+0.0912	−0.0144	—
			MID-PANEL		
n_1.....	−0.0128	+0.0848	+0.1504	−0.0240	+0.0016
n_2.....	−0.0016	+0.0144	+0.2224	−0.0416	+0.0064
n_3.....	+0.0064	−0.0336	+0.2544	−0.0336	+0.0064
n_4.....	+0.0064	−0.0416	+0.2224	+0.0144	−0.0016
n_5.....	+0.0016	−0.0240	+0.1504	+0.0848	−0.0128
			LAST NEXT-TO-END PANEL		
n_1.....	−0.0144	+0.0912	+0.1408	−0.0176	—
n_2.....	−0.0080	+0.0400	+0.1840	−0.0160	—
n_3.....	+0.0000	−0.0080	+0.2160	−0.0080	—
n_4.....	+0.0080	−0.0480	+0.2320	+0.0080	—
n_5.....	+0.0144	−0.0752	+0.2272	+0.0336	—
			LAST END-PANEL		
n_1.....	+0.0176	−0.0848	+0.1968	+0.0704	—
n_2.....	+0.0160	−0.0720	+0.1360	+0.1200	—
n_3.....	+0.0080	−0.0320	+0.0400	+0.1840	—
n_4.....	−0.0080	+0.0400	−0.0960	+0.2640	—
n_5.....	−0.0336	+0.1488	−0.2768	+0.3616	—

Source: United Nations, Methods for Population Projections by Age and Sex. ST/SOA/ Series A, Population Studies, No. 25

Table 3.11

Using Sprague Multipliers to Apportion the 15-19 Population Into Single Years

	M_1	M_2	M_3	M_4	M_5	
	5-9	10-14	15-19	20-14	25-29	
	447,017[a]	389,292	337,539	311,422	306,415	Single Years
n_1	- .0128	.0848	.1504	- .0240	.0016	15 = 71,070
n_2	- .0016	.0144	.2224	- .0416	.0064	16 = 68,966
n_3	.0064	- .0336	.2544	- .0336	.0064	17 = 67,148
n_4	.0064	- .0416	.2224	.0144	- .0016	18 = 65,730
n_5	.0016	- .0240	.1504	.0848	- .0128	19 = 64,625
						337,539

APPENDIX B

ADDITIONAL INFORMATION ON LIFE TABLES

Fortunately, educational planners usually do not have to build life tables. Demographers do this and life tables are available for most countries, or for states or provinces within countries.

Table 3.12 is a state life table. In the 3rd column it shows 100,000 people in the age group, 0-1. This column is symbolized "lx", the number alive at the start of an age group interval. The table uses 100,000 because this is the base to which rates will be expressed (just as percents use the base 100, and fertility uses the base 1000).

The number 100,000 is then multiplied by the probability of dying in that age interval (symbolized as "qx" and shown in the 2nd column). This gives the number that die in the age interval 0-1, which is symbolized "dx" (4th column). The dx figure is 1,725 in the table. The 1,725 is then subtracted from 100,000 to give lx, the number alive at the start of the next age interval, or age group, 1-2. This is procedure is repeated through all the ages of the table.

Table 3.13 and the explanatory information below it give more information on life tables.

Table 3.12

Life Table for the Total Population: Massachusetts, 1969-71

AGE IN YEARS	PROPORTION DYING	OF 100,000 BORN ALIVE		STATIONARY POPULATION		AVERAGE REMAINING LIFETIME
PERIOD OF LIFE BETWEEN TWO EXACT AGES STATED (1)	PROPORTION OF PERSONS ALIVE AT BEGINNING OF YEAR OF AGE DYING DURING YEAR (2)	NUMBER LIVING AT BEGINNING OF YEAR OF AGE (3)	NUMBER DYING DURING YEAR OF AGE (4)	IN YEAR OF AGE (5)	IN THIS YEAR OF AGE AND ALL SUBSEQUENT YEARS (6)	AVERAGE NUMBER OF YEARS OF LIFE REMAINING AT BEGINNING OF YEAR OF AGE (7)
SYMBOL	q_x	l_x	d_x	L_x	T_x	$\mathring{e}_x$
0-1	0.01725	100,000	1,725	98,441	7,187,497	71.87
1-2	.00085	98,275	84	98,233	7,089,056	72.09
2-3	.00072	98,191	71	98,155	6,990,823	71.15
3-4	.00053	98,120	52	98,094	6,892,668	70.28
4-5	.00040	98,068	40	98,044	6,789,933	69.24
5-6	.00041	98,028	40	98,008	6,691,889	68.27
6-7	.00038	97,988	37	97,961	6,593,889	67.30
7-8	.00035	97,963	34	97,926	6,495,928	66.32
8-9	.00031	97,909	31	97,874	6,398,002	65.35
9-10	.00027	97,878	26	97,865	6,300,128	64.37
10-11	.00024	97,852	24	97,839	6,202,263	63.38
11-12	.00023	97,828	23	97,817	6,104,424	62.40
12-13	.00027	97,805	26	97,791	6,006,607	61.41
13-14	.00035	97,779	35	97,762	5,908,796	60.43
14-15	.00048	97,744	47	97,720	5,811,034	59.45
15-16	.00063	97,697	61	97,667	5,713,314	58.48
16-17	.00077	97,636	75	97,599	5,615,647	57.52
17-18	.00088	97,561	86	97,518	5,518,048	56.56
18-19	.00095	97,475	93	97,429	5,420,530	55.61
19-20	.00096	97,382	95	97,334	5,323,101	54.66
20-21	.00100	97,287	97	97,239	5,225,767	53.71
21-22	.00105	97,190	100	97,140	5,128,528	52.77
22-23	.00104	97,090	102	97,039	5,031,388	51.82
23-24	.00104	96,988	100	96,938	4,934,349	50.88
24-25	.00103	96,888	100	96,838	4,837,411	49.93
25-26	.00100	96,788	97	96,739	4,740,573	48.98
26-27	.00098	96,691	95	96,643	4,643,834	48.03
27-28	.00097	96,597	94	96,549	4,547,189	47.07
28-29	.00101	96,503	97	96,455	4,450,640	46.12
29-30	.00107	96,406	104	96,354	4,354,185	45.17
30-31	.00117	96,302	112	96,246	4,257,831	44.21
31-32	.00128	96,190	124	96,128	4,161,585	43.26
32-33	.00138	96,066	133	95,999	4,065,457	42.32
33-34	.00146	95,933	139	95,864	3,969,458	41.38
34-35	.00152	95,794	146	95,721	3,873,594	40.44
35-36	.00157	95,648	150	95,573	3,777,873	39.50
36-37	.00167	95,498	159	95,418	3,682,300	38.56
37-38	.00184	95,339	176	95,251	3,586,882	37.62
38-39	.00211	95,163	201	95,062	3,491,631	36.69
39-40	.00244	94,962	232	94,846	3,396,569	35.77
40-41	.00281	94,730	266	94,597	3,301,723	34.85
41-42	.00315	94,464	298	94,316	3,207,126	33.95
42-43	.00344	94,166	326	94,003	3,112,810	33.06
43-44	.00373	93,840	350	93,665	3,018,807	32.17
44-45	.00399	93,490	373	93,303	2,925,142	31.29
45-46	.00425	93,117	396	92,920	2,831,839	30.41
46-47	.00456	92,721	423	92,509	2,738,919	29.54
47-48	.00496	92,298	457	92,070	2,646,410	28.67
48-49	.00547	91,841	505	91,589	2,554,340	27.81
49-50	.00605	91,336	554	91,060	2,462,751	26.96
50-51	.00677	90,782	615	90,474	2,371,691	26.13
51-52	.00748	90,167	675	89,830	2,281,217	25.30
52-53	.00824	89,492	737	89,124	2,191,387	24.49
53-54	.00904	88,755	802	88,354	2,102,263	23.69
54-55	.00988	87,953	869	87,519	2,013,909	22.90

Source: National Center for Health Statistics. U.S. Decennial Life Tables for 1969-70, Vol. II, No. 22, "Massachusetts State Life Tables: 1969-71."

Table 3.13

Some Functions of an Abridged Life

Age [illegible] (x)	[illegible] death rate (1,000 m_x)	[illegible] death rate (1,000 m_x)	[illegible] (l_x)	[illegible] (L_x)	[illegible] (P_x)	T_x	Expectancy ^0e_x
0.....	71.1	63.0	100,000	467,715	0.9062	(not shown)	62.1
1-4...	3.1	12.0	93,700				
5-9.....	1.4	7.0	92,576	461,260	0.9937	--	65.2
10-14..	1.1	5.5	91,928	458,375	0.9933	--	62.0
15-19..	1.6	8.0	91,422	455,282	0.9971	--	57.4
20-24...	2.4	11.9	90,691	450,755	0.9879	--	52.7
25-29 ..	2.5	12.4	89,611	445,278	0.9859	--	
30-34..	3.2	15.9	88,500	438,982	0.9872	--	
35-39....	4.0	19.8	87,093	431,155	0.9761	--	
40-44...	5.7	28.1	85,369	420,818	0.9626	--	
45-49....	9.6	47.0	82,970	405,100	0.9490	--	
50-54.....	11.3	55.1	79,070	384,458	0.9284	--	
55-59	16.6	89.1	74,713	356,922	0.8918	--	
60-64..	27.5	129.1	[illegible]	[illegible]	0.8412	--	
65-69..	42.5	192.9	59,270	267,768	0.7676	--	
70-74...	65.2	251.1	47,837	205,545	0.6649	--	
75-79..	103.4	410.0	34,381	136,642	0.5226	--	
80-84...	173.4	592.2	20,284	71,190	0.3075	--	
85 and over.	240.9	1,000.0	8,272	31,406			

Source: United Nations, Methods for Population Projections by Age and Sex. ST/SDA/ Series A. Population Studies, No. 25

Explanation of life table columns and how computed

1. The first column x indicates the age group.
2. The second column (1000 m_x) is the age-specific death rate.
3. The third column is the probability of dying before reaching upper limit of age group. (Explanation Follows).
4. The fourth column l_x is the number of survivors to exact age. It begins with 100,000 and is reduced by the number dying in each age group.
5. The fifth column L_x is survivors within the age group (average number of years lived). It can be computed (except for the first age group) by averaging successive l values.
6. The column ^xP_x shows the proportion in age group surviving to next age group, and is obtained by dividing successive L_x values.
7. The T_x column would be obtained by cumulating up from the lowest L_x through each successive age interval.
8. The last column ^0e_x is the number of years (expected) of life remaining at the beginning of the year of age. From 0 it is commonly called life expectation at birth. It is obtained by dividing the T_x value by l_x value.

We can use values from Table 3.13 to illustrate how the q values (probability of dying in a given age interval) are generated from the m values

(age-specific death rates), using the curve-fitting equation (1) below. Once the q values are obtained, the rest of the columns of a life table can be computed.

${}_nq_x$ = probability of dying within n years after age x

$${}_nq_x = 1-e^{-n\,{}_nm_x - an^3\,{}_nm_x^2},$$

where $a = 0.008$

$n = 5$

${}_nm_x = 18.6/1000$

$= .0186$

Thus, $1000q_x =$

$$1000\left[1-e^{-5(.0186)-(.008)(5)^3(.0186)^2}\right]$$

$$= 1000\left[1-e^{-.0833}\right]$$

$$= 1000(.0891) = 89.1$$

Exercise

THE POPEX MODEL

The terms "model" and "template" mean the same thing when referring to small, component, spreadsheet models. Popex model, was designed in Lotus by Cassidy, based on a model by Davis(1980). Popex will project total population, school-age population and the school-age entrant pool, using a components method of projection, described in Chapter Two. It works best for planning in the national education systems of LDC's, when external migration is not significant at national level. Popex does permit the inclusion of migration data, but it is not used in this example (although it is shown, it is not used).

```
-----------------------------------------------------
|                    POPEX2                          |
|                 Version: Test 1                    |
|                                                    |
|   A cohort-based population projection template    |
|       for use with LOTUS 1-2-3, Version 2.0        |
|                 Requires 512K  RAM.                |
|                                                    |
|              The BRIDGES Project                   |
|       Harvard Graduate School of Education         |
|                                                    |
|                   T. Cassidy                       |
|                    R. Davis                        |
|                                                    |
|                   May 1986                         |
-----------------------------------------------------
```

[Press the Tab-key once]

The template is designed in a series of screen views shown in the Template Map in Annex A. In general the structure of the design is a series of Introductions and Descriptions, which we have extracted from the actual computer screens and shown on these next few pages. They are the title screen (shown above), the Introduction and the Description Screens and the Popex Macro Menus. These different screens run along the top of the spreadsheet from A1 to Y20 and can be accessed by a stroke on the Tab Key, which takes you across from left to right (Shift/Tab takes you back from right to left).

Here is what the INTRODUCTION screen will look like on your computer:

INTRODUCTION

Before attempting to use POPEX2 you should read the five screens to the right of this one. To move from screen to screen simply press the Tab-key. To move backwards press the Tab-key while holding the Shift-key down.

This template has been structured around a set of Command Menus with the intention of making it easier to move about the template, input data, view results tables and graphs, and print selected results tables. These are presented in the 4th screen. It is also possible to move about the template freely using the Tab, Shift & Tab, PgUp, and PgDn keys. To make this free-form movement easier template components have been layed out at even screen distances from one another starting at the HOME position.

A general familiarity with LOTUS 1-2-3, particularly with the form and use of 1-2-3 Menus, will make use of this template easier. If you are a new comer to 1-2-3, locate the ESC-key on the keyboard, it can be used to escape from most any unexpected situation. If ever ESC shouldn't work, hold the CTRL-key down, press the Break-key, release both and then press the ESC-key.

Description Screens (There are three)

DESCRIPTION (Screen 1 of 3)
This covers the purpose and use of Popex.

DESCRIPTION (Screen 1 of 3)

POPEX2 is a program that:

1. Projects individual cohort and total population fifteen years out from a base year in three 5-year periods.

2. Interpolates age-sex specific populations for ages 0-19 for the base year and each five year period.

3. Calculates the number of 4, 5, 6, and 7 year olds for use in school enrollment projections.

DESCRIPTION (Screen 2 of 3)
-This covers the basic demographics of Popex:

a) Basic balance equation: Base Year Population and the three demographic components;

b) Description of the main components: Births or Fertility; Deaths (Mortality); and Migration (which is not inserted in this exercise).

'DESCRIPTION' (Screen 2 of 3)

Populations are projected using the following method:

FUTURE POP = BASE YR POP + BIRTHS - DEATHS +/- MIGRATION

Births, meaning survivors of birth, are a function of the number of women of child-bearing age multiplied by a set of fertility rates which are then multiplied by survival ratios obtained from the UN Model Life Tables.

Deaths are calculated using appropriate survival ratios for specific age-sex cohorts obtained from the UN Model Life Tables.

Populations, projected by age-sex cohorts, are distributed to specific ages using Sprague Multipliers.

DESCRIPTION (Screen 3 of 3)
-This is just a note on Migration as a kind of disclaimer.

'DESCRIPTION' (Screen 3 of 3)

'NOTE: Migration is not a fully integrated component of this version of POPEX2. Thus, this template is of limited use in situations where national level migration is a significant factor.

POPEX MENUS

-This description (exact replica below) which you will find in the top rows of your template covers the Popex Menus:

1. The Components Menu;
2. The Input Menu;
3. The Population Tables (or Projection) Menu;
4. The Graphs Menu;
5. The Print Menu;
6. The Help Menu;

Menus are shown as they appear on Popex menu line, Annex B.

You will notice that each one of these menus are selected by simultaneously pressing the ALT key and a letter corresponding to the function you desire. The letter "C" with the ALT key invokes the Componets Menu. The letter "I" stands for, and is an abreviation for, Input. These menus theoretically may be called up at any time, from almost any point inside the template.

NOTE: If the Alt-key combinations do not work at any point later on, try the following sequence to re-initialize the menu and to get them engaged again: Press ESC, then "/" (do not use the quotes). Then try ESC, F5 (the "goto" key) and the ALT-key combination of your choice.

POPEX2 MENUS

Reviewing the various components of POPEX2, inputting data, viewing graphs, printing, and accessing this and other help screens are most easily accomplished using one of several selection menus. Menus are invoked by pressing a specified key while holding the ALT-key down.

The menus available and the keys used to invoke them are:

1. The 'Components Menu'	ALT-key & C-key
2. The 'Input Menu'	ALT-key & I-key
3. The 'Population Tables Menu'	ALT-key & T-key
4. The 'Graphs Menu'	ALT-key & G-key
5. The 'Print Menu'	ALT-key & P-key
6. The 'Help Menu'	ALT-key & H-key

*For example of sensitivity test go to cell Q41

Use the Space Bar to make selections within menus and the Return-key to invoke your selection.

NOTE: You may want to copy the above table on paper.

The Components and the Input Menus

The Components Menu accesses the fertility rates assumed and the survival rates from:

Model Life Tables,
Births and Survivors,
Migration,
UN Model Life Tables, and
Sprague Multiplier panels.

The Input Menu does NOT directly use the Life Tables and Multipliers. When you need a figure from the tables for the input menu, they are provided for your reference, so you have a convenient place right in the template to go look them up. You can even use the "copy" command to bring them to your input location.

Components Menu

BASE YR POP	FERTILITY	MORTALITY	BIRTHS	MIGRATION	LIFE TABLE	SPRAGUE MULTIPLIER

INPUT MENU

BASE YR POP	FERTILITY	MORTALITY	BIRTHS	MIGRATION

Note: See Annex B for subheading explanations of the menus' essential data, since the component data is input to a components' projection.

The "Components" menu and the "Input" menu access the same essential data since the component data is input to a components' "projection" (the section where the actual calculation is done).

The following pages contain explanations of the selections from the "Components" and "Inputs" menus:

BASE YR POP

Selecting BASE YR POP on the Input or Components menus takes the user to the base year population screen in which the population is arrayed by sex and five year age groups, as the one shown for the Popex example shown below. Both the "Components" menu and the "Input" menu begin with the Base Year Population, which is located directly under the first screen of the template (as shown in the Template Map --A21 to D44). The Base Year Population in this Popex example exercise is printed below as it would be called up on your screen when you press ENTER while BASE YR POP is selected (highlighted by the cursor).

BASE YEAR POPULATION 1980

AGE-GROUP	MALE	FEMALE	TOTAL
0-4	475291	459374	934665
5-9	402120	392669	794789
10-14	314449	340742	655191
15-19	302190	315684	617874
20-24	299000	307654	606654
25-29	269765	274567	544332
30-34	248768	256787	505555
35-39	234543	243543	478086
40-44	217657	229800	447457
45-49	188765	201324	390089
50-54	245673	261460	507138
55-59	215678	236789	452467
60-64	191345	201222	392567
65-69	175673	189789	365467
70-74	98890	122345	221235
75-79	64599	85678	150277
80+	28970	46789	75759
TOTAL	3973386	4166216	8139602

FERTILITY

Selecting FERTILITY on either of the two menus (Alt-C or ALT-I) calls up (or takes the user to) the two tables below. The first assumes constant fertility over the forecast, and the second assumes the rates (age-specific) decline over the three five-year forecast periods.

FERTLITY (rates per 1000 women)

Which hypothesis? (1 or 2) 1 (Enter data below in the table of your choice.)

Hypothesis 1 - Unchanging Fertility (Enter data for all age cohorts in the Base year only.)

AGE	Base Yr 1980	1985	1990	1995
15-19	85.0	85.0	85.0	85.0
20-24	222.0	222.0	222.0	222.0
25-29	250.0	250.0	250.0	250.0
30-34	198.7	198.7	198.7	198.7
35-39	139.0	139.0	139.0	139.0
40-44	64.9	64.9	64.9	64.9
45-49	15.8	15.8	15.8	15.8

Hypothesis 2 - Declining Fertility (Enter data for all age cohorts, all years.)

AGE	1980	1985	1990	1995
15-19	84.4	81.3	78.4	75.7
20-24	223.3	214.3	206.7	199.4
25-29	260.4	251.3	242.2	233.6
30-34	198.7	191.5	184.9	178.3
35-39	139.0	134.0	129.2	124.6
40-44	64.9	62.6	60.4	58.3
45-49	15.8	15.2	14.7	14.1

Mortality shows survival rate levels from UN Model Life Tables, improving over the forecast. Note: The Fertility and Mortality components shown on the previous page, permit choices based on different assumptions. This is what "projections" are. In Popex, choice points are back-lighted to indicate that they can be changed on the template as input, whereas other labels and data arrays (model tables) are cell protected (you are locked out) against accidental change.

MORTALITY

Survival Ratios (Source - UN Model Life Tables)

PERIOD		UN MORTALITY LEVEL
1980 TO	1985	80
1985 TO	1990	85
1990 TO	1995	90

NOTE: To view the UN Model Life Tables use the 'Components Menu' (ALT-key & C-key).

Births and Survivors are shown below for the three periods. Births are estimated on the basis of age specific fertility rates multiplied by the average number of women (between base and forecast periods) in the age group. (Male/Female Birth Fractions are applied, and birth survival rates to each year, 0-4 are applied.) This gives the new numbers in the 0-4 age group for the next forecast period. Note that the 653780 females gets calculated as the 0-4 group for year 1990 in the Projection Table for 1990 (below). (Recall: all the other age groups from 5 years old and above were derived as survivors from the previous age group in the previous period.) Births and Survivors are projected for the three periods of the exercise, 1980-1985; 1985-1990; and 1990-1995. This also shows that the process is simply iterative, and every five years new rates are input, and the procedures is the same.

BIRTHS AND SURVIVORS

1980 - 1985 BIRTH FRACTION (% FEMALES) = 0.4911

FERTILE AGE-GROUP	AVG # of WMN in GRP	5-YR FERTILITY	TOT BIRTHS	TOT FEMALE BIRTHS	SURVIVAL RATIO	1985 TOT FEMALE SURVIVE	TOT MALE BIRTHS	SURVIVAL RATIO	1985 TOT MALE SURVIVE	1985 TOT M & F SURVIVE
15-19	326611	0.4250	138810	68170	0.9208	62771	70640	0.9070	64071	126841
20-24	309538	1.1100	343587	168736	0.9208	155372	174852	0.9070	158590	313962
25-29	288633	1.2500	360792	177185	0.9208	163152	183607	0.9070	166532	329684
30-34	263288	0.9935	261577	128460	0.9208	118286	133116	0.9070	120737	239023
35-39	247738	0.6950	172178	84557	0.9208	77860	87621	0.9070	79473	157332
40-44	234029	0.3245	75942	37295	0.9208	34342	38647	0.9070	35053	69394
45-49	212425	0.0790	16782	8241	0.9208	7589	8540	0.9070	7746	15335
						619377			632201	1251572

FERTILE AGE-GROUP	AVG # of WMN in GRP	5-YR FERTILITY	TOT BIRTHS	TOT FEMALE BIRTHS	SURVIVAL RATIO	1990 TOT FEMALE SURVIVE	TOT MALE BIRTHS	SURVIVAL RATIO	1990 TOT MALE SURVIVE	1990 TOT M & F SURVIVE
15-19	361955	0.4250	153831	75546	0.9380	70863	78285	0.9262	72507	143370
20-24	322641	1.1100	358132	175878	0.9380	164974	182253	0.9262	168803	333777
25-29	305021	1.2500	381277	187245	0.9380	175636	194032	0.9262	179712	355348
30-34	284095	0.9935	282249	138613	0.9380	130019	143637	0.9262	133036	263055
35-39	258730	0.6950	179818	88308	0.9380	82833	91509	0.9262	84756	167589
40-44	242752	0.3245	78773	38686	0.9380	36287	40088	0.9262	37129	73416
45-49	228021	0.0790	18014	8847	0.9380	8298	9167	0.9262	8491	16789
						668916			684435	1353344

FERTILE AGE-GROUP	AVG # of WMN in GRP	5-YR FERTILITY	TOT BIRTHS	TOT FEMALE BIRTHS	SURVIVAL RATIO	1995 TOT FEMALE SURVIVE	TOT MALE BIRTHS	SURVIVAL RATIO	1995 TOT MALE SURVIVE	1995 TOT M & F SURVIVE
15-19	413865	0.4250	175893	86381	0.9380	81025	89512	0.9262	82906	163931
20-24	358454	1.1100	397885	195401	0.9380	183286	202484	0.9262	187540	370827
25-29	318865	1.2500	398581	195743	0.9380	183607	202838	0.9262	187369	371476
30-34	301089	0.9935	299133	146904	0.9380	137796	152229	0.9262	140994	278790
35-39	279995	0.6950	194597	95566	0.9380	89641	99030	0.9262	91722	181363
40-44	254276	0.3245	82513	40522	0.9380	38010	41991	0.9262	38892	76901
45-49	237261	0.0790	18744	9205	0.9380	8634	9539	0.9262	8835	17469
						722007			738757	1460757

(Note: From the main menu screen try simply using the "cursor-down" key and the "cursor-right" key to explore this section. In any case, be sure if you do select BIRTHS from the menu to use the "cursor-right" key to see the whole page, or you will be missing out on a portion of this.)

Migration

Migration in this exercise is simply a place-holder component to indicate that in some countries' external migration is significant, and data are available to deal with it as a component. In most countries, neither of these conditions holds. In the US, and other countries, migration is put in as a constant and it is age and sex specific (as shown below for Females, included as an example in this exercise).

MIGRATION - FEMALES

AGE-GROUP	1980 TO 1985	1985 TO 1990	1990 TO 1995
0-4	1	1	1
5-9	2	2	2
10-14	3	3	3
15-19	4	4	4
20-24	5	5	5
25-29	6	6	6
30-34	7	7	7
35-39	8	8	8
40-44	9	9	9
45-49	10	10	10
50-54	11	11	11
55-59	12	12	12
60-64	13	13	13
65-69	14	14	14
70-74	15	15	15
75-79	16	16	16
80+	17	17	17
TOTAL	153	153	153

Model Life Tables

On the following page are Survival Rates for Males and Females, excerpted from UN Model Life Tables and shown as they are provided in your template. Other model tables not provided are the Coale Demeny Regional Tables (1966). The age-specific survival rates are tabled in 7 levels based on increasing levels of life expectancy at birth.

Sprague Multipliers

The Sprague Multiplier panels, also accessible from the Components menu, spreads five year age groupings into single years. A replica, here, can be found five pages over.

LIFE TABLES (Survival ratios)

LEVEL 60	LEVEL 65	LEVEL 70	LEVEL 75	LEVEL 80	LEVEL 85	LEVEL 90	LEVEL 95	LEVEL 100	LEVEL 105	LEVEL 110	LEVEL 115
(AA 50.0)	(AA 52.5)	(AA 55.0)	(AA 57.6)	(AA 60.4)	(AA 63.2)	(AA 65.8)	(AA 68.2)	(AA 70.2)	(AA 71.7)	(AA 73.0)	(AA 73.9)
0.8406	0.8557	0.8703	0.8877	0.9070	0.9262	0.9438	0.9580	0.9678	0.9744	0.9788	0.9817
0.9445	0.9518	0.9584	0.9648	0.9708	0.9765	0.9818	0.9867	0.9908	0.9963	0.9954	0.9966
0.9834	0.9856	0.9876	0.9893	0.9909	0.9924	0.9937	0.9949	0.9963	0.9974	0.9982	0.9986
0.9827	0.9848	0.9867	0.9884	0.9900	0.9914	0.9928	0.9941	0.9955	0.9968	0.9977	0.9982
0.9741	0.9772	0.9798	0.9824	0.9848	0.9871	0.9893	0.9913	0.9935	0.9953	0.9965	0.9974
0.9688	0.9725	0.9758	0.9789	0.9819	0.9847	0.9873	0.9898	0.9924	0.9944	0.9957	0.9966
0.9673	0.9712	0.9748	0.9781	0.9811	0.9840	0.9866	0.9891	0.9916	0.9935	0.9948	0.9956
0.9644	0.9686	0.9725	0.9760	0.9792	0.9822	0.9849	0.9874	0.9897	0.9917	0.9930	0.9939
0.9580	0.9628	0.9671	0.9711	0.9746	0.9779	0.9808	0.9835	0.9860	0.9881	0.9897	0.9907
0.9467	0.9523	0.9575	0.9617	0.9658	0.9696	0.9728	0.9759	0.9786	0.9812	0.9835	0.9851
0.9294	0.9359	0.9419	0.9469	0.9516	0.9558	0.9596	0.9632	0.9664	0.9696	0.9727	0.9752
0.9045	0.9121	0.9189	0.9250	0.9306	0.9356	0.9400	0.9441	0.9481	0.9521	0.9558	0.9594
0.8678	0.8767	0.8847	0.8919	0.8985	0.9044	0.9098	0.9148	0.9198	0.9249	0.9296	0.9341
0.8141	0.8244	0.8338	0.8423	0.8502	0.8573	0.8639	0.8702	0.8765	0.8830	0.8890	0.8945
0.7367	0.7487	0.7598	0.7698	0.7794	0.7881	0.7963	0.8042	0.8119	0.8197	0.8269	0.8335
0.6334	0.6469	0.6594	0.6708	0.6816	0.6917	0.7012	0.7102	0.7192	0.7283	0.7372	0.7454
0.5049	0.5193	0.5326	0.5449	0.5566	0.5675	0.5779	0.5878	0.5976	0.6072	0.6176	0.6280
0.2988	0.3091	0.3186	0.3272	0.3352	0.3426	0.3493	0.3556	0.3613	0.3666	0.3721	0.3775
0.8594	0.8739	0.8882	0.9036	0.9208	0.9380	0.9535	0.9660	0.9744	0.9801	0.9838	0.9864
0.9463	0.9537	0.9607	0.9669	0.9731	0.9791	0.9844	0.9892	0.9925	0.9948	0.9963	0.9973
0.9828	0.9852	0.9874	0.9895	0.9914	0.9932	0.9948	0.9962	0.9972	0.9981	0.9986	0.9990
0.9819	0.9843	0.9865	0.9886	0.9906	0.9925	0.9941	0.9956	0.9967	0.9976	0.9982	0.9987
0.9745	0.9778	0.9809	0.9838	0.9865	0.9891	0.9914	0.9935	0.9952	0.9965	0.9974	0.9980
0.9694	0.9735	0.9771	0.9807	0.9839	0.9869	0.9896	0.9920	0.9940	0.9956	0.9966	0.9973
0.9674	0.9718	0.9756	0.9792	0.9826	0.9858	0.9884	0.9909	0.9929	0.9945	0.9956	0.9963
0.9654	0.9700	0.9739	0.9777	0.9811	0.9842	0.9868	0.9893	0.9913	0.9929	0.9941	0.9948
0.9622	0.9669	0.9710	0.9749	0.9783	0.9814	0.9841	0.9866	0.9886	0.9903	0.9915	0.9923
0.9556	0.9606	0.9650	0.9691	0.9727	0.9759	0.9788	0.9814	0.9835	0.9853	0.9868	0.9879
0.9437	0.9493	0.9543	0.9589	0.9631	0.9668	0.9700	0.9731	0.9754	0.9775	0.9792	0.9806
0.9252	0.9318	0.9377	0.9433	0.9482	0.9529	0.9570	0.9609	0.9637	0.9661	0.9680	0.9695
0.8949	0.9030	0.9104	0.9175	0.9238	0.9298	0.9351	0.9402	0.9437	0.9466	0.9491	0.9510
0.8459	0.8559	0.8653	0.8743	0.8825	0.8903	0.8971	0.9036	0.9083	0.9122	0.9161	0.9196
0.7710	0.7832	0.7948	0.8059	0.8162	0.8259	0.8345	0.8429	0.8490	0.8549	0.8609	0.8668
0.6662	0.6804	0.6940	0.7073	0.7197	0.7316	0.7422	0.7525	0.7602	0.7677	0.7764	0.7852
0.5369	0.5523	0.5672	0.5819	0.5955	0.6088	0.6207	0.6323	0.6410	0.6493	0.6592	0.6697
0.3211	0.3313	0.3406	0.3495	0.3576	0.3652	0.3719	0.3781	0.3828	0.3869	0.3913	0.3957

The Projection Results

The results of using the projection template can be called up by a menu which looks as follows: (You may get back to the main menu with the F5 "goto" command and the cell designation AQ1)

Population Tables (sub-menu, called up with ALT-T)

POP-F	POP-M	POP-TOT (No Migrat)	POP-TOT (w Migrat)	Age/Sex	New Entrants

The projection results are:

1) The Female Population, projected to 1985, 1990 and 1995;
2) Male Population, same years;
3) Total Population without Migration;
4) Total Population (with Migration "dummied-in");
5) Age/Sex all periods;
6) New Entrants which are age-sex single age for ages 4, 5, 6, 7, which cover most possible entry ages in most countries of the world. This last is, of course, the point of the model, as far as educational planners go.

POPULATION PROJECTION - FEMALES

AGE-GROUP	BASE 1980	SURVIVAL RATIO	FEMALE POP 1985	SURVIVAL RATIO	FEMALE POP 1990	SURVIVAL RATIO	FEMALE POP 1995
0-4	459374	0.9731	619377	0.9791	668916	0.9844	722007
5-9	392669	0.9914	447017	0.9932	606432	0.9948	658481
10-14	340742	0.9906	389292	0.9925	443977	0.9941	603279
15-19	315684	0.9865	337539	0.9891	386372	0.9914	441358
20-24	307654	0.9839	311422	0.9869	333860	0.9896	383050
25-29	274567	0.9826	302701	0.9858	307343	0.9884	330388
30-34	256787	0.9811	269790	0.9842	298402	0.9868	303777
35-39	243543	0.9783	251934	0.9814	265527	0.9841	294464
40-44	229800	0.9727	238258	0.9759	247248	0.9788	261305
45-49	201324	0.9631	223526	0.9668	232516	0.9700	242006
50-54	261460	0.9482	193895	0.9529	216105	0.9570	225541
55-59	236789	0.9238	247916	0.9298	184763	0.9351	206813
60-64	201222	0.8825	218746	0.8903	230513	0.8971	172772
65-69	189789	0.8162	177578	0.8259	194749	0.8345	206793
70-74	122345	0.7197	154906	0.7316	146662	0.7422	162518
75-79	85678	0.5955	88052	0.6088	113329	0.6207	108853
80+	46789	0.3576	67753	0.3652	78349	0.3719	99481
TOTAL	4166216		4539702		4955064		5422884

POPULATION PROJECTION - MALES

AGE-GROUP	BASE 1980	SURVIVAL RATIO	MALE POP 1985	SURVIVAL RATIO	MALE POP 1990	SURVIVAL RATIO	MALE POP 1995
0-4	475291	0.9708	632201	0.9765	684435	0.9818	738757
5-9	402120	0.9909	461413	0.9924	617344	0.9937	671978
10-14	314449	0.9900	398461	0.9914	457906	0.9928	613455
15-19	302190	0.9848	311305	0.9871	395034	0.9893	454609
20-24	299000	0.9819	297597	0.9847	307289	0.9873	390807
25-29	269765	0.9811	293588	0.9840	293043	0.9866	303386
30-34	248768	0.9792	264666	0.9822	288891	0.9849	289117
35-39	234543	0.9746	243594	0.9779	259955	0.9808	284528
40-44	217657	0.9658	228586	0.9696	238210	0.9728	254964
45-49	188765	0.9516	210213	0.9558	221637	0.9596	231731
50-54	245678	0.9306	179629	0.9356	200922	0.9400	212682
55-59	215678	0.8985	228628	0.9044	168061	0.9098	188866
60-64	191345	0.8502	193787	0.8573	206771	0.8639	152902
65-69	175678	0.7794	162682	0.7881	166133	0.7963	178630
70-74	98890	0.6816	136923	0.6917	128209	0.7012	132292
75-79	64599	0.5566	67403	0.5675	94710	0.5779	89900
80+	28970	0.3352	45667	0.3426	53897	0.3493	73559
TOTAL	3973386		4356341		4782447		5262164

POPULATION PROJECTION - TOTAL (FEMALE & MALE) WITH NO MIGRAT

AGE-GROUP	BASE 1980	1985	1990	1995
0-4	934665	1251578	1353351	1460764
5-9	794789	908429	1223777	1330459
10-14	655191	787753	901883	1216734
15-19	617874	648844	781406	895967
20-24	606654	609019	641149	773857
25-29	544332	596289	600386	633774
30-34	505555	534456	587293	592894
35-39	478086	495527	525482	578992
40-44	447457	466844	485458	516269
45-49	390089	433740	454153	473737
50-54	507138	373524	417027	438223
55-59	452467	476544	352823	395679
60-64	392567	412532	437284	325673
65-69	365467	340260	360883	385422
70-74	221235	291829	274871	294810
75-79	150277	155455	208039	198753
80+	75759	113420	132246	173040
TOTAL POP	8139602	8896043	9737510	10685048

NEW ENTRANTS (found in the sub-menu called up with ALT-T) are obtained by splitting the five year age group 0-4 into single years, 0-1, 1-2, 2-3, 3-4, 4-5 and then splitting the 5-9 into 5-6, 6-7 etc. Below are shown the results for the 5, 6 and 7 year old males and females.

NEW ENTRANTS: AGE 5	BASE YEAR 1980	1985	1990	1995
FEMALE	83571	97959	132434	134855
MALE	88036	100593	134512	137910
TOTAL	171608	198553	266945	272765

NEW ENTRANTS: AGE 6	BASE YEAR 1980	1985	1990	1995
FEMALE	80944	92677	127824	133117
MALE	84418	95470	129880	136000
TOTAL	165362	188147	257704	269117

NEW ENTRANTS: AGE 7	BASE YEAR 1980	1985	1990	1995
FEMALE	78416	88486	122086	131630
MALE	80540	91420	124208	134330
TOTAL	158956	179906	246294	265959

SPRAGUE MULTIPLIERS

The single year age groups are obtained by using Sprague Multipliers to split the 0-4, 5-9, 10-14, & 15-19 age groupings into single years. This is shown below and is found by ALT-C then Sprague (From location BQ22 go right).

INTERPOLATION OF NUMBERS OF MALES AGED 0-14 BY SINGLE YEARS OF AGE IN YEAR 1980 USING SPRAGUE MULTIPLIERS

0-4	5-9	10-14	15-19	20-24	SUM	
475291	402120	314449	302190	299000		MALES
0.3616	-0.2768	0.1488	-0.0336		97195	
0.2640	-0.0960	0.0400	-0.0080		97034	AGE 1
0.1840	0.0400	-0.0320	0.0080		95893	AGE 2
0.1200	0.1360	-0.0720	0.0160		93918	AGE 3
0.0704	0.1968	-0.0848	0.0176		91251	AGE 4
				check	475291	(AGES 0-4)
0.0336	0.2272	-0.0752	0.0144		88036	AGE 5
0.0080	0.2320	-0.0480	0.0080		84418	AGE 6
-0.0080	0.2160	-0.0080	0.0000		80540	AGE 7
-0.0160	0.1840	0.0400	-0.0080		76546	AGE 8
-0.0176	0.1408	0.0912	-0.0144		72580	AGE 9
				check	402120	(AGES 5-9)
-0.0128	0.0848	0.1504	-0.0240	0.0016	68535	AGE 10
-0.0016	0.0144	0.2224	-0.0416	0.0064	64306	AGE 11
0.0064	-0.0336	0.2544	-0.0336	0.0064	61286	AGE 12
0.0064	-0.0416	0.2224	0.0144	-0.0016	60120	AGE 13
0.0016	-0.0240	0.1504	0.0848	-0.0128	60201	AGE 14
				check	314449	AGES(10-14)

INTERPOLATION OF NUMBERS OF MALES AGED 15-19 BY SINGLE YEARS OF AGE IN 1985 USING SPRAGUE MULTIPLIERS

AGES	5-9	10-14	15-19	20-24	25-29	SUM	
	402120	314449	302190	299000	269765		MALES
15	-0.0128	0.0848	0.1504	-0.0240	0.0016	60223	AGE 15
16	-0.0016	0.0144	0.2224	-0.0416	0.0064	60380	AGE 16
17	0.0064	-0.0336	0.2544	-0.0336	0.0064	60565	AGE 17
18	0.0064	-0.0416	0.2224	0.0144	-0.0016	60574	AGE 18
19	0.0016	-0.0240	0.1504	0.0848	-0.0128	60448	AGE 19
					check	302190	(AGES 15-19)

Summary Comment on POPEX

Annex A shows the Template Map of Popex, with the screens identified by spreadsheet location. A template map simply shows view of the spreadsheet as if you "zoomed" away from the close-up view of the template in order to see the whole thing at once--just as an aerial photo would provide for a large area of earth terrain. Its value is to show not only location identifiers but also the positions in relation to each other.

Annex B shows the menus designed with the various macros, which are located in the MACRO screens on the Template Map.

Annex C shows the design of the Macros, so that the template can be structurally customized to fit the needs of any user. Any given user, of course, can restructure the tables and change the data to fit a given planning and projection situation. That is the whole point of template models.

Once the model is implemented and the projections made for any given planning situation, a number of spreadsheet features can be used to assist planners:

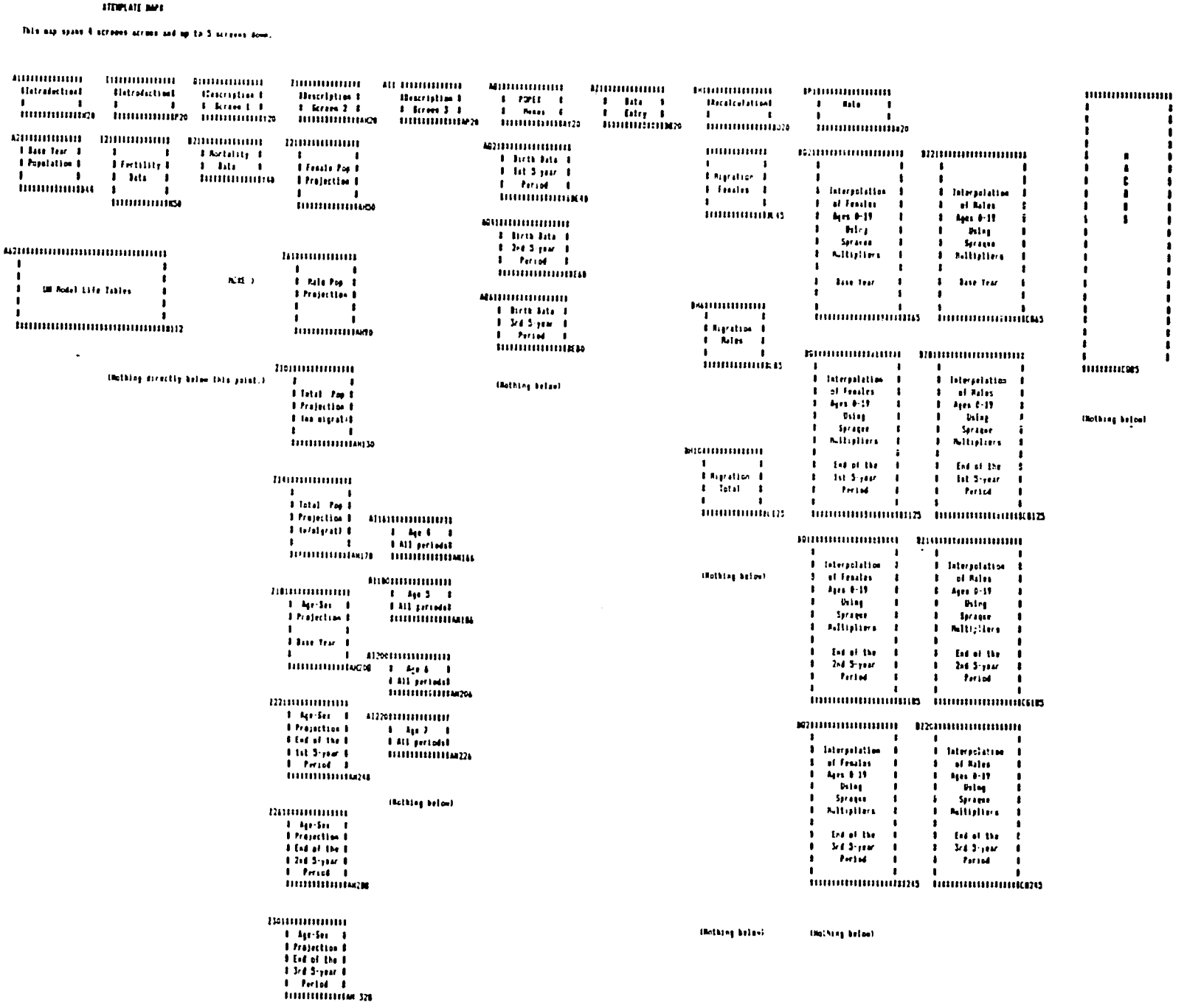
1) "What-If" and hypothesis testing can be done by changing any of the data values, parameter values or even structural characteristics to fit the specific country and school system. A planner could, for example, experiment with changed hypotheses about the demographic components (increased births) and trace the consequences for new entrants in future years.

2) Analysts can also develop sensitivity tables (in Popex some versions have demonstration Sensitivity table areas). In sensitivity testing, the analyst can change one or more values of variables, and assess the degree to which other variable values are sensitive to these changes.

One feature of Popex should be noted, and the instructions do alert users to this: Automatic Recalculation of the spreadsheet values has been turned off, and the Recalc is on manual, as a lighted area at the foot of the spreadsheet shows the user. The reason for this is that the spreadsheet is quite large and changes take time to make in a large system. Another reason is to remind the user that mindless changes--without prior systematic thought, and policy and program reference--are of small use in planning.

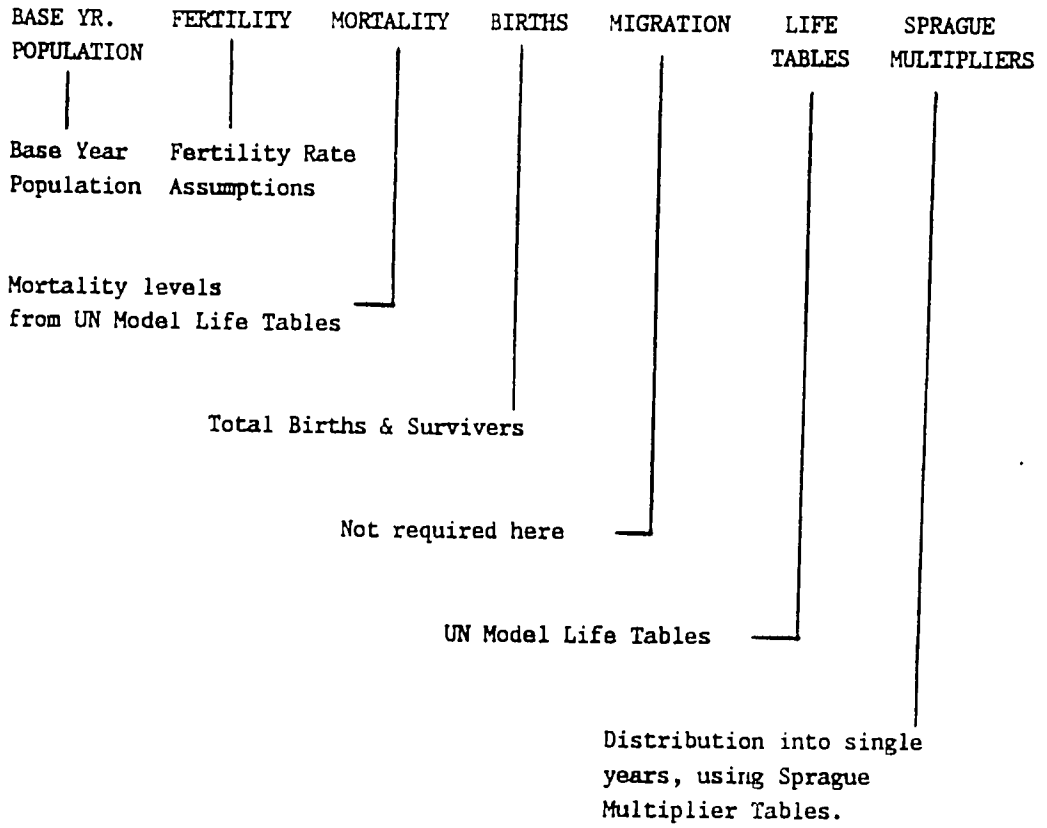
Changes in hypotheses, and the building of scenarios out of sets of changes and hypotheses should be carefully thought out beforehand, in terms of the real life system that the

template model represents. When that is done and the changes are introduced, and the changes in the results observed, then the planner is doing the kind of analysis which is useful and valuable. Spreadsheet models are suited to this work of experimenting with different input values.

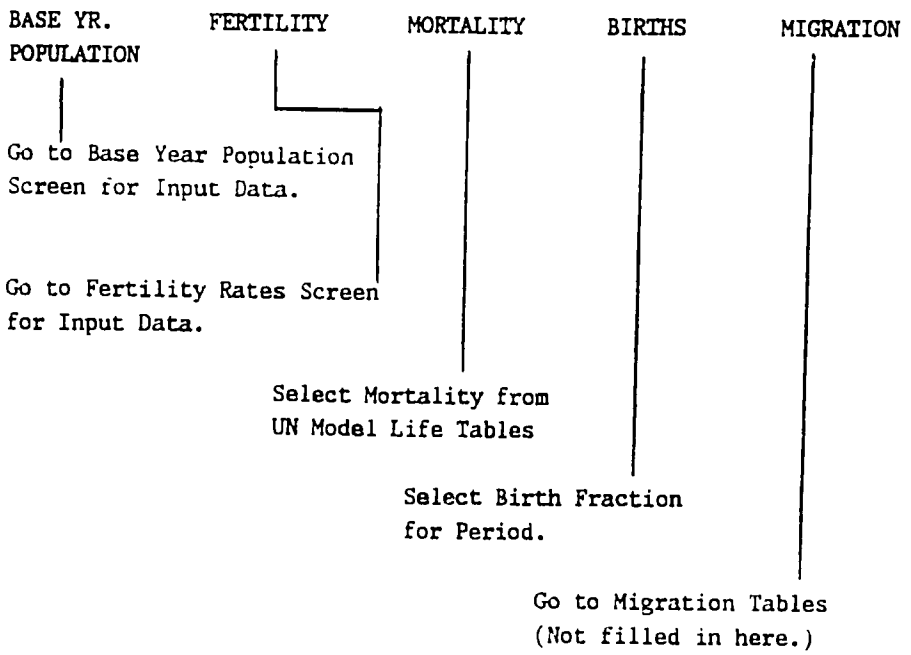

TEMPLATE MAPS
This map spans 4 screens across and up to 5 screens down.
Introduction
Description Screen 1
Description Screen 2
Description Screen 3
Base Year Population
Fertility Data
Mortality Data
UN Model Life Tables
Female Pop Projection
Male Pop Projection
Total Pop Projection
Total Pop Projection (w/algrat)
Age-Sex Projection
Projection End of the 1st 5-year Period
Projection End of the 2nd 5-year Period
Projection End of the 3rd 5-year Period
Birth Data 1st 5 year Period
Birth Data 2nd 5 year Period
Birth Data 3rd 5 year Period
Recalculation
Migration Females
Migration Males
Migration Total
Interpolation of Females Ages 0-19 Using Sprague Multipliers
Interpolation of Males Ages 0-19 Using Sprague Multipliers
Base Year
End of the 1st 5-year Period
End of the 2nd 5-year Period
End of the 3rd 5-year Period
Age 0
Age 5
Age 6
Age 7
All periods
(Nothing below)

ANNEX B
(Screen Menus)

Components Menu:



Input Menu:



Population Tables Menu: (projection exercise results)

POP-F POP-M POP-TOT(no mig.) POP-TOP(migration) AGE/SEX NEW ENTRANT

Population Projections Females

Population Projections Males

Total Male/Female no migration

Total Male/Female with migration

Age/Sex Projections for all periods.

New Entrants Projected. Ages 4, 5, 6, 7.

Graphics Menu:

LINE GRAPHS

Line Graphs of Population Projections

BAR GRAPHS

Bar Graphs of Population Projections

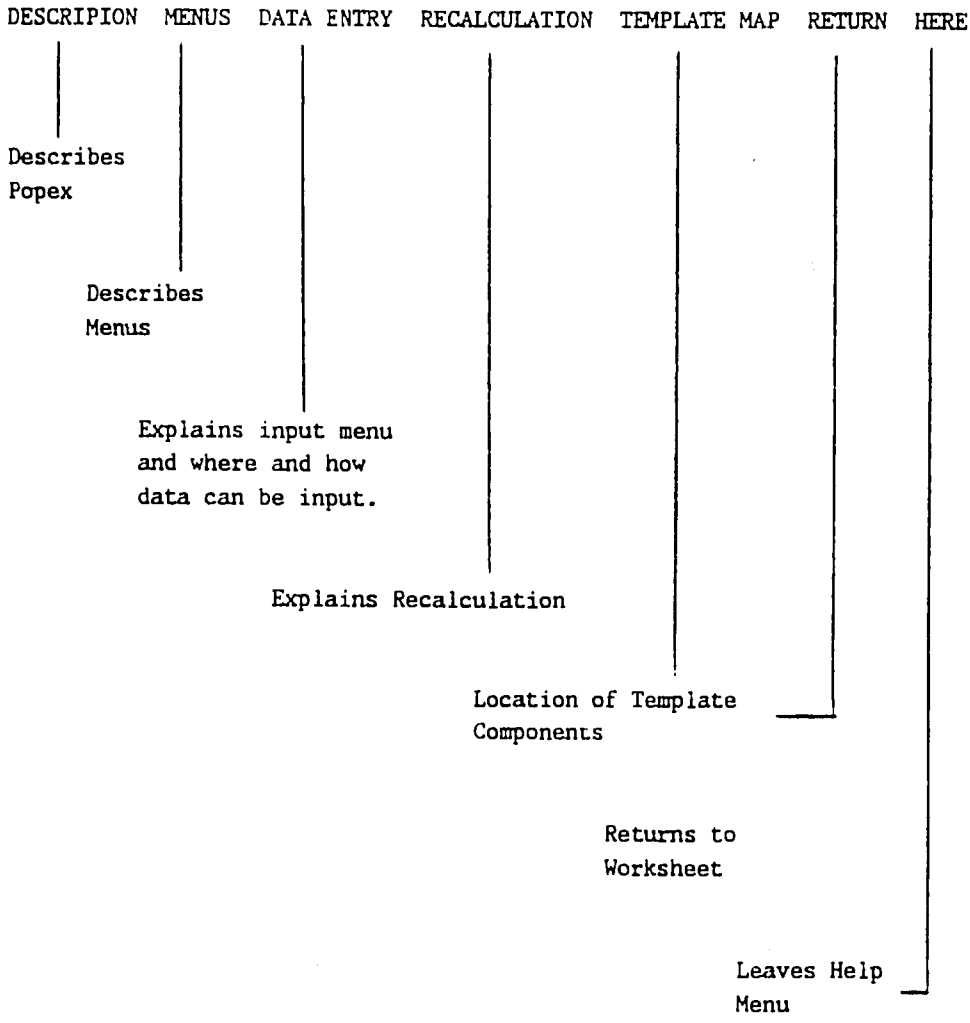
Print Menu:

POP TABLES

AGE-SEX TABLES

(Note: will print when menu selection is invoked.

Help Menu:


DESCRIPION
MENUS
DATA ENTRY
RECALCULATION
TEMPLATE MAP
RETURN
HERE
Describes Popex
Describes Menus
Explains input menu and where and how data can be input.
Explains Recalculation
Location of Template Components
Returns to Worksheet
Leaves Help Menu

ANNEX C

```
\c      {menubranch locate}

Locate  BASE YR POP          FERTILITY            MORTALITY            BIRTHS               MIGRATION    LIFE TABLES  SPRAGUE
        Base Year Population. Fertility rate assumptions. Mortality levels useTotal births and sur{Not a requireUN Model Life The calculation of age - s
        {GOTO}BASE'          {GOTO}FERTIL'        {GOTO}MORTAL'        {BRANCH B&S}         {BRANCH MIGRTN{GOTO}LIFTAB' {BRANCH \s}

                                                                                                   distributions using Sprague Multipliers.
\T      {menubranch popproj}

Popproj POP-F                POP-M                POP-TOT(no migrat)   POP-TOT(w/migrat)    AGE-SEX PROJ NEW ENTRANTS
        Population projections foPopulation projections for aTotal(female + male)Total(female + male)Projections for all periods by age & sex, ages 0-19.
        {GOTO}FEMALES'       {GOTO}MALES'         {GOTO}NOMIGRAT'      {GOTO}W/MIGRAT'      {BRANCH \a}  {BRANCH \E}

\A      {menubranch agesex}

AgeSex  BASE YR              1ST PROJ PERIOD      2ND PROJ PERIOD      3RD PROJ PERIOD      NEW ENTRANTS
        Base year population agesEnd of 1st 5-year period. pEnd of 2nd 5-year peEnd of 3rd 5-year peMenu for selection of appropriate new entrant age grou
        {GOTO}YBASE'         {GOTO}ONE'           {GOTO}TWO'           {GOTO}THREE'         {BRANCH \E}

\S      {menubranch sprague}

Sprague BASE YR-F            BASE YR-M            1ST-F                1ST-M                2ND-F        2ND-M        3RD-F        3RD-M
        Interpolation of the numbInterpolation of the number The number of femaleThe number of males The number of The number of The number of The number of
        {GOTO}BF'            {GOTO}BM'            {GOTO}OF'            {GOTO}OM'            {GOTO}TF'    {GOTO}TM'    {GOTO}THRF'  {GOTO}THRM'

\G      {menubranch graph}                                                males ages 0-19 at the end of the third projection period.

Graph   LINE GRAPHS          BAR GRAPHS
        Line graphs of populationBar graphs of population projections.
        {BRANCH LINE}        {BRANCH BAR}

\H      {paneloff}/mdHERE'
        /mdHERE''{panelon}
Menu    {menubranch helpmenu}

Helpmenu DESCRIPTION         MENUS                DATA ENTRY           RECALCULATION        TEMPLATE MAP RETURN       HERE
        A description of POPEX2. Menus that allow you to moveExplanation of allowInformation on menusLocations of tReturn to the Stop here.
        {GOTO}HELP1'         {GOTO}HELP2'         {GOTO}HELP3'         {GOTO}RECALC'        {GOTO}MAP1'  {GOTO}HERE'  {QUIT}'
        {BRANCH DESCRIP}     {BRANCH MENU}        {BRANCH MENU}        {BRANCH MENU}        {BRANCH MENU} {QUIT}'

                                                                                   leave the Help Menu, do not return to previous location.
\E      {menubranch entrants}

Entrants AGE 4               AGE 5                AGE 6                AGE 7
        Use if new entrants to thUse if new entrants to the sUse if new entrants Use if new entrants to the system are normally
        {GOTO}AGE4'          {GOTO}AGE5'          {GOTO}AGE6'          {GOTO}AGE7'                        7 years old.

NEWLINE {menubranch new}

        AGE 4                AGE 5                AGE 6                AGE 7
        New entrants age 4, projeNew entrants age 5, projecteNew entrants age 6, New entrants age 7, projected on a line graph.
        /gnuAAL'q'           /gnuA5L'q'           /gnuA6L'q'           /gnuA7L'q'
```

NEBAR {menubranch neb}

AGE 4	AGE 5	AGE 6	AGE 7
New entrants age 4, proje	New entrants age 5, projecte	New entrants age 6,	New entrants age 7, projected on a bar graph.
/gnuA4b"q"	/gnuA5b"q"	/gnuA6b"q"	/gnuA7b"q"

MIGRTN {menubranch migrat}

FEMALE	MALE	TOTAL
Female migration for all	Male migration for all perio	Total migration for all periods.
{GOTO}MIGF"	{GOTO}MIGM"	{GOTO}MIGT"

DESCRIP {menubranch screens}

DESCRIPTION-Pg1	DESCRIPTION-Pg2	DESCRIPTION-Pg3	HELP	RETURN	HERE
Move to Description Scree	Move to Description Screen	Move to Description	Return to the Help M	Return to the	Stop here.
{GOTO}HELP1"	{GOTO}HELP1A"	{GOTO}HELP1B"	{BRANCH MENU}	{GOTO}HERE"	
{BRANCH DESCRIP}	{BRANCH DESCRIP}	{BRANCH DESCRIP}		{QUIT}	

leave the Help Menu. Go not return to previous location.

B&S {menubranch births}

1ST PROJ PERIOD	2ND PROJ PERIOD	3RD PROJ PERIOD
Births and survivors duri	Births and survivors during	Births and survivors during the 3rd five-year projection period.
{GOTO}B&S1"	{GOTO}B&S2"	{GOTO}B&S3"

Line {menubranch pl}

TOTAL POP	FEMALES	MALES	FEM & MALE	NEW ENTRANTS
Total projected populatio	Projected female populations	Projected male popul	Comparison of female	New entrants to school.
/gnuTOTPOP_LINE"q"	/gnuFEMLINE"q"	/gnuMALELINE"q"	/gnuM&FLINE"q"	{BRANCH NELINE}

represented on a line graph.

Bar {menubranch pb}

TOTAL POP	FEMALES	MALES	FEM & MALE	NEW ENTRANTS
Total projected populatio	Projected female populations	Projected male popul	Comparison of female	New entrants to school,
/gnuTOTPOP_BAR"q"	/gnuFEMBAR"q"	/gnuMALEBAR"q"	/gnuM&FBAR"q"	{BRANCH NEBAR}

represented on a line graph.

\P {Menubranch print}

POP TABLES	AGE-SEX TABLES
{BRANCH POPTABL}	{BRANCH A-STABL}

Poptabl {menubranch poptab}

Poptab TOTAL POP
Total population projections by cohorts, all projection periods.
{paneloff}{windowsoff}
/ppr{bs}z101..ag124~agg
{panelon}{windowson}

A-stabl {menubranch A-stab}

BASE YR	1ST PROJ PERIOD	2ND PROJ PERIOD	3RD PROJ PERIOD	NEW ENTRANTS
Ages 0-19 in the base yea	Ages 0-19 at the end of the	Ages 0-19 at the end		Ages 0-19 at the end
{paneloff}{windowsoff}	{paneloff}{windowsoff}	{paneloff}{windowsof	{paneloff}{windowsof	{BRANCH PRINTNEW}
/ppr{bs}Z181..AF208~agg	/ppr{bs}Z221..AF248~agg	/ppr{bs}Z261..AF288~	/ppr{bs}Z301..AF328~agg	
{panelon}{windowson}	{panelon}{windowson}	{panelon}{windowson}	{panelon}{windowson}	

Printnew {menubranch pnew}

AGE 4	AGE 5	AGE 6	AGE 7
{paneloff}{windowsoff}	{paneloff}{windowsoff}	{paneloff}{windowsof	{paneloff}{windowsoff}
/ppr{bs}AI160..AN166~agg	/ppr{bs}AI180..AN186~agg	/ppr{bs}AI200..AN206	/ppr{bs}AI220..AN226~agg
{panelon}{windowson}	{panelon}{windowson}	{panelon}{windowson}	{panelon}{windowson}

\I {menubranch input}

Input	BASE YR POP	FERTILITY	MORTALITY	BIRTHS	MIGRATION
	Go to Base Year Populatio	Go to Fertility Rates Screen	Select Mortality Lev	Select period for en	Go to migration tables.
	{GOTO}BASE~{RIGHT 3}{?}	{GOTO}FERTIL~{RIGHT 3}{DOWN	{GOTO}MORTAL~{RIGHT	{BRANCH B&SINPUT}	{BRANCH MIGRTN}

{Note: Migration is not a required input.}

{menubranch B&SINP}

1ST PROJ PERIOD	2ND PROJ PERIOD	3RD PROJ PERIOD
Births and survivors duri	Births and survivors during	Births and survivors during the 3rd five-year projection period.
{GOTO}B&S1~{RIGHT 8}{DOWN	{GOTO}B&S2~{RIGHT 8}{DOWN 2}	{GOTO}B&S3~{RIGHT 8}{DOWN 2}

Chapter 4

ENROLLMENT FORECASTING

4.0 INTRODUCTION

The future development of a school system will depend on the numbers and kinds of students entering the various grades, levels, and programs in the system. On the basis of future enrollment estimates, decisions are made about program coverage and the need for teachers, materials, and facilities. Once these decisions are made, costs and future budgets are estimated. Enrollment forecasts--when compared to population estimates and population forecasts--are also important for measuring overall performance of educational systems.

In manpower planning, economists first forecast demand. Enrollment projections provide the basis for supply estimates. By comparing demand and supply estimates, it is possible to estimate the number of trained workers of various types needed in a nation. Thus, enrollment forecasting is critically important to a nation both in terms of planning for education and for economic stability.

Enrollment forecasts are based in part on population estimates and populations projections: Population estimates show the school age population now. Population projections (or forecasts) show the future school age population. Knowing the present and future school age population is essential in planning because:

1. The number eligible to enter the first level of the system is a basic component of school enrollment forecasts.

2. Comparison of the age eligible population to actual school enrollments provides two important indicators of system performance. Comparison of children of school entrance age with actual first grade enrollments yields an intake rate. A broader comparison of eligible people actually served by the schools yields a measure of overall coverage. Both may be used to measure the extent to which school systems are attaining their goals and targets. They also show how much remains to be done to attain goals and targets.

targets.

Enrollment forecasting is based on three components, (each of which will be described in the next sections):

- The school entrant pool (the number of the population eligible to enter the first level of the system). Planners are interested in the past, present, and future entrant pool for a system.

- The number of entrants (the percent of age eligible children actually admitted into the first grade). This is calculated by multiplying the entrant pool by the intake rate (the percent of age eligible children actually admitted to the first grade).

- Flow rates of enrollments through the system. These rates are determined by promotion (p), repetition (r), and dropout (d) in the school system.

4.1 DETERMINING THE ENTRANT POOL: COMPONENT 1

Population forecasting supplies the data for the first component of enrollment forecasting. Using the results of a population forecast (as described in Chapter 3), planners can estimate the age eligible entrant pool. The entrant pool includes all children who reach school entrance age in a given year (usually 5 or 6 years of age).

4.2 DETERMINING THE NUMBER OF ENTRANTS: COMPONENT 2

Once the entrant pool number is known, the number of entrants into the system can be forecast by fixing an intake rate for admission. The intake rate is the percent of children that will actually be admitted into the system. Multiplying the entrant pool number by the intake rate will give the number of entrants. For example, a population forecast might give the following figures:

Total school age children in Year X = 22 million
Entrant pool for first grade in Year Y = 4.1 million

Then, if the intake rate stated in a national plan is set at 80%, the number of expected entrants is calculated by multiplying the number in the entrant pool by the intake rate:

4.1 million X .80 = 3.28 million Year Y.

This component might be considered a simple matter of multiplication. However, the question arises: How is the intake rate determined?

Intake rates can be set at whatever target percent the writers of plan or policy goals choose to make them. In addition to setting intake rates for the general population, officials and planners may target particular groups in the population for special attention. For example, intake rates for females, rural females, tribal groups, urban poor, or groups with low enrollment rates may be set at higher percentages than those for the general population.

It is good to remember, however, that goals are statements of intent. (They often are also political statements.) An intake rate indicates the percent of the entrant pool that policy makers and officials desire to see admitted to the system. But, as all experienced planners know, attainment of goals is often limited by a lack of resources required to provide needed inputs into the system, such as teachers, classroom space, instructional materials, etc.

Reasonable estimates of resource requirements and costs can only be made on the basis of forecasts of entrants and enrollments. When enrollment projections are known, planners can then estimate how much it will cost to provide the resources for that number. They must then compare these estimates of with estimates of revenues that will be available to the system. On this basis realistic plan targets can be set.

Enrollment forecasting, then, might sometimes lead to adjusting goals toward more realistic and attainable intake rate targets over a period of years. For example, a system might desire an intake rate of close to 100% for the general population, but the following procedure may actually be more realistic:

4.1 million	X	.80	=	3.28 million	Year Y
4.3 million	X	.85	=	3.66 million	Year Y + 5
4.7 million	X	.90	=	4.23 million	Year Y + 10

(Entrant Pool) (Intake Rate) (Number of Entrants)

Adjusting goals and targets does not mean abandoning them. It means moving toward the targets realistically within the limits of resources and over a realistic period of time. The problem is, as the simple data above show,

of time. The problem is, as the simple data above show,

each year the target gets higher for two reasons: (1) The entrant age population increases and (2) entrants unserved in past years may also build up a backlog and thus increase the number to be admitted. These factors must also be considered in setting intake rates. (But then, nobody ever said educational planning was supposed to be easy!)

Once the entrant pool has been estimated (first component) and the number of entrants for a system is determined (second component), it is time to consider the third component of enrollment forecasting: Flow rates.

4.3 DETERMINING FLOW RATES: COMPONENT 3

The flow through a school system from first to final grade is determined by three rates which are called flow rates. These rates include the (1) the number promoted, (2) the number of those who repeat a grade, and (3) the number of those who drop out of the system.

1. Promotion Rate (p)

Promotion rates reflect the number of students promoted from one grade to the next. For example, promotion rates can be calculated for the number who are promoted from Grade 1 to Grade 2, from Grade 2 to Grade 3, and on to Grade 6, (or the final grade in the level). Promotion rates are also determined for flows from level to level. For example, a promotion rate might be determined for the number who go from the last grade of the elementary level to the first grade of the intermediate level. Promotion rates are symbolized by the letter p.

Promotion rates are determined by dividing the number of students promoted by the total number enrolled.

Example:

100 children are enrolled in Grade 1 in 1985.
80 of those children are promoted to Grade 2 in 1986.

Grade 1 promotion rate (p) for 1985 = 80%: 80/100 = .80.

2. Repetition Rate (r)

The repetition rate is the proportion of students who repeat a grade (stay in the same grade for another year). Repetition rates are symbolized by the letter r.

Example:

100 children are enrolled in Grade 1 in 1985.
15 of those children repeat Grade 1 in 1986.

Grade 1 repetition rate (r) for 1985 = 15%: 15/100 = .15

3. Dropout Rate (d)

The dropout rate is the proportion of students who drop out of the system at a given grade level. The dropout rate includes students who leave the system because of death, migration, or any other reason. Dropout rates are symbolized with the letter d.

Example:

100 children are enrolled in Grade 1 in 1985.
5 children drop out before entering Grade 2 in 1986.

Grade 1 dropout rate (d) for 1985 = 5%: 5/100 = .05

It is important to recognize that p, r, and d rates are calculated for each grade level and each year. Thus, there may be one set of p, r, and d rates for Grade 1 in 1985 and another set of rates for the same grade in 1986.

Summing the p, r, and d rates calculated in the three examples on the previous page will show another truth about flow rates:

Flow rates for Grade 1 (1985)

Promotion rate (r)	= .80
Repetition rate (r)	= .15
Drop out rate (d)	= .05
	1.00

Notice that the rates add to 1: (p + r + d = 1). The rates add up to 1, or to 100 percent, because all students must be promoted, repeat, or drop out.

The previous sections have covered the three components of enrollment forecasting: (1) Determining the entrant pool, (2) Determining the number of entrants, and (3) Determining the flow rates. These components are the inputs to the basic model for enrollment forecasting which will be discussed in the next section.

4.4 ENROLLMENT FORECASTING: THE BASIC MODEL

In the best of all worlds, planners have complete data provided by a pupil accounting system that follows each student through the education system. Promotion, repetition, and dropout rates can be computed exactly for each grade level and then used by planners in accurate forecasts. If complete and accurate data are available, enrollment forecasts are simple.

The list of seven steps below will serve as an example of enrollment forecasting when complete data are available on the three components (including all p, r, and d rates). Keep in mind that enrollment forecasting begins with a current (or base) year and then makes estimates from that year for the next year (or years).

Step 1
Forecast the entrant pool--the population that is age eligible to enter Grade 1. In most cases, this will be the number of six year-olds.

Step 2
Multiply the total number of age eligible children by the intake rate to determine the number of entrants into Grade 1.

Step 3
Multiply the current Grade 1 enrollment by the Grade 1 repetition rate (r) and add this number to the entrants. This gives the total enrollment in Grade 1 for the next year.

Step 4
Multiply the current Grade 1 enrollment by the Grade 1 dropout rate (d). This number goes out of the system and should be subtracted from the total moving into Grade 2.

Step 5
Multiply the current Grade 1 enrollment by the Grade 1 promotion rate (p). Move these enrollments to Grade 1, Year 2 of the forecast.

Step 6
Multiply the Grade 2 enrollment by the Grade 2 repetition rate. Add this to the number promoted from Grade 1. This will give the total Grade 2 enrollment. Multiply this number by the Grade 2 dropout rate and subract this number before moving the enrollment into Grade 3 of the next year.

Step 7
Follow this procedure for Grade 3, Grade 4, and all succeeding grades and years in the forecast. Like population forecasts (and most forecast methods), the process is repetitive.

In the real world, of course, planners almost never have data or information in the right form for use. Few systems have complete pupil accounting systems, (though such systems will undoubtably become more common in the future). In typical situations, planners may have the exact numbers enrolled in each grade*, but not the numbers of those promoted, repeaters or dropouts (the p, r, and d rates). Yet, the model described on the previous page assumed that flow rates were available. In cases where complete data are not available, planners must use other models for enrollment forecasting.

The section which begins on the next page will present several models for estimating flow rates under less favorable conditions. When flow rates are not routinely calculated by the system, planners select one of these models based upon the completeness (or incompleteness) of the data collected by the system. Some of the models make it possible to make enrollment estimates even when very little data is available. However, it must be emphasized that accurate forecasting is dependent upon accurate and complete data. Nevertheless, the models which follow can be used to estimate enrollment until pupil accounting systems provide the necessary data.

*In some systems, enrollments are arranged by both age and grade. See Appendix A at the end of this chapter for a method developed by Schiefelbein for age-grade data.

method developed by Schiefelbein for age-grade data.

4.5 ENROLLMENT FORECASTING WITH INCOMPLETE DATA

In many cases, planners have only aggregate data--total enrollments by grade--with which to work. Working with aggregate data, rather than data which includes p, r, and d rates, calls for different models and methods. Depending on the data available, a planner would select one of the following models:

1. Aggregate Data for Only One Year
2. Aggregate Data for Multiple Years
3. Aggregate Data for Multiple Years with Repeaters

1. Aggregate Data for Only One Year

In the very worst case, enrollment data would be available for only one year. For example, Table 4.1 gives data for the 1985/86 school year for six grades, rounded off in 1000's.

Table 4.1

Enrollment Data, Grades 1-6 (1985/86)

Grade 1	Grade 2	Grade 3	Grade 4	Grade 5	Grade 6
1,138.7	1,124.6	1,019.0	988.2	873.6	858.8

With only this limited data with which to work, each successive grade would be treated as a surviving proportion of the grade before. This proportion is called a <u>transition rate</u> (or <u>survival rate</u>). For example, from Grade 1 to Grade 2 the transition rate would be calculated by dividing the Grade 2 enrollment by the enrollment of grade 1:

$$1,124.6 / 1,138.7 = .9876$$

The Grade 2 to Grade 3 transition rate would be .9061, and so on. These figures would be taken as the proportion surviving from one grade into the next.

However, there are problems in calculating transition rates this way. Look at this example:

1976 Enrollment, Grade 1: 766,088
1976 Enrollment, Grade 2: 820,319

1985 Grade 1 Tranistion Rate: 1.071
(820,319/766,088 = 1.071)

From your reading earlier in the chapter, can you explain why 1.071 cannot possible be a survival rate? (An explanation will be given in the next section.)

Fortunately, this worst case is rarely found. Most planners have at least the enrollment data from multiple years to work with, as the next section will discuss.

2. Aggregate Data for Multiple Years

Table 4.2 shows enrollments for Egypt by Grades 1 to 6 for the years 1976/77 through 1985/86. (This table actually the complete version of Table 3.1 which listed only the 1976/77 data.)

Table 4.2

Egyptian Enrollment, Grades 1-6

1976-1986

ANALYSIS OF EGYPTIAN ENROLLMENT DATA,GRADES 1-6,YEARS 1976-1986

Enrollment

Year\Gr	1	2	3	4	5	6
1976-77	766,088	820,319	691,171	636,563	592,466	635,529
1977-78	786,005	837,675	710,200	743,532	531,786	602,147
1978-79	793,572	852,252	718,246	766,454	629,224	527,876
1979-80	823,914	853,967	739,267	772,257	650,621	594,531
1980-81	861,141	877,926	753,968	789,550	666,482	598,991
1981-82	922,100	911,776	791,363	796,285	702,154	624,736
1982-83	969,136	961,318	844,950	823,842	732,367	704,995
1983-84	1,039,423	1,010,643	901,971	877,986	767,168	752,388
1984-85	1,077,873	1,082,437	950,388	939,065	817,524	813,241
1985-86	1,138,689	1,124,628	1,019,027	988,199	873,561	858,746

The Grade-to-Grade Survival or Transition Rate Model just discussed makes much more sense when multiple year enrollment data by grades is available. With these data, transition rates are computed down the diagonal of the table. Table 4.3 on the next page shows a portion of Table 4.2 (1976 to 1983). The diagonal line shows the pattern for calculating more accurate transition rates.

Table 4.3

Multiple Year, Multiple Grade Diagonal Analysis

(Enrollment in 1000's)

Year/ Grade	1	2	3	4	5	6
1976/77	766.1	820.3	etc.	.	.	.
1977/78	786.0	838.0	718.3	etc.	.	.
1978/79	793.6	852.3	718.3	767.0	etc.	.
1979/80	824.0	854.0	739.3	772.3	651.0	etc.
1980/81	861.1	877.9	754.0	790.0	667.0	.
1981/82	922.1	912.0	791.4	796.3	702.2	625.0
1982/83	969.2	961.3	845.0	824.0	732.4	705.0

Transition rates can now be computed based on actual enrollment figures for a span of years. For example, to determine the transition rates for the group entering first grade in 1976, each successive grade would be divided by the one before it. For example, the calculations for the diagonal indicated on Table 4.3 would be as follows:

Grade 1 to Grade 2: 838.0/766.1 = 1.0939
Grade 2 to Grade 3: 718.3/838.0 = .8572
Grade 3 to Grade 4: 772.3/718.3 = 1.0752
Grade 4 to Grade 5: 667.0/772.3 = .8637
Grade 5 to Grade 6: 625.0/667.0 = .9370

With this model it is again possible to have transition rates greater than 1. What is the explanation for this? It would be impossible to have transition rates higher than 1 if they were real survival fractions from an original cohort, as with the survival P's in the population forecasting models discussed in Chapter 3.

The difference is that the diagonal on Table 4.3 does not represent a true cohort. A cohort is a group that shares a time-based characteristic such as the same year of birth, same year of entry into school, etc. When the diagonal model is used with enrollments that enter and leave the system, the group in Grade 2, Year 2 are not necessarily survivors from Grade 1 Year 1. Different students have come in and gone out of the group. Does the 1.0939 transition rate from Grade 1 to Grade 2 indicate a large number of new entrants to Grade 2? Does it indicate a large number of repeaters? There is no way to know for sure from these data. (See Appendix B at the end of the chapter for additional information on Cohort Analysis.)

Despite the weaknesses of the model, additional analyses can be done by tracking other groups. Here are two examples:

1. Two different years can be compared

Grade 1 1976/77 to Grade 2 in 1977/78 = 838.0/766.1 = 1.0939

Grade 1 1977/78 to Grade 2 in 1978/79 = 852.3/786.0 = 1.0843

Presumably, the rates from one grade to another (for example Grade 1 to Grade 2) would be fairly consistent over time.

2. Transition rates for the same grade for multiple years can be averaged to give more reliable figures

1976/77		1.0939
1977/78		1.0843
other years	+	.
		2.1782 / 2 = 1.0891

1.0891 = Average Transition Rate for Grade 1 to Grade 2.

Even with the strangeness of having survival fractions greater than 1, the diagonal model can be applied to estimate transition rates. The output of the model comes close enough to reality to be usable as a start in enrollment forecasting. Later, as the country's pupil accounting system improves, better data can be used. (Planners often must work with approximate data and improve it over time.)

The third model to be discussed uses data on repeaters along with multiple year enrollment data. This section begins on the next page.

3. Aggregate Data, Multiple Years with Repeaters

The next best case for a planner is to have data showing total enrollment figures for multiple grades and years, but also data on those students who have repeated a grade. (The data on repeaters is shown in parentheses in Table 4.4 below.) With only the addition of the information on repeaters, planners can calculate promotion, repetition, and dropout rates for all grade to grade transitions. The diagonal analysis model is again used here.

Table 4.4

Total Enrollments and Repeaters by Grade and Year

(Repeaters Shown in Parentheses)

Grade\Year	1982	1983	1984	1985	1986	1987
1	50225	55210 (16114)	59876 (17012)	59811 (18104)	66010 (18487)	67363
2	37125	40283 (8276)	42582 (8305)	43394 (8877)	47008 (9258)	51794
3	28383	30975 (5719)	33091 (5699)	35712 (6576)	38662 (6485)	41806
4	20611	22621 (3268)	24038 (3265)	25761 (3712)	29028 (3579)	32531
5	13867	15598 (1444)	16986 (1595)	18321 (1818)	20352 (1745)	23643 (1631)
6	9886	11328 (394)	12355 (328)	13451 (385)	14938 (374)	17062 (341)

Working with enrollment and repeater figures, planners can calculate the other flow rates needed for enrollment forecasting. The method for calculating each rate will be described on the next page using Grade 1 in 1982 as an example. (Once the rates on individual years and grades are computed, they can be compared, averaged, and so on.)

Grade 1 (1982)
Example

Promotion rate:

First subtract the number of children repeating Grade 2 in 1983 from those who were promoted from Grade 1 to Grade 2 at the end of the 1982/83 school year. Then divide this number by the total enrollment in Grade 1 in 1982 to determine the proportion of students who were promoted:

40283 - 8276 = 32007 / 50225 = .6373 (p)

Repetition rate:

Dividing the number of students who repeated Grade 1 in 1983 by the total enrollment in Grade 1 in 1982 will give the repetition rate:

16114 / 50225 = .3208 (r)

Dropout rate:

First, two things must be subtracted from the total enrollment in Grade 1 in 1982:

(1) The number of children repeating Grade 2 in 1983 from those who were promoted from Grade 1 to Grade 2 at the end of the 1982/83 school year (32007) and

(2) The number of repeaters in Grade 2 in 1983 16114).

Having eliminated those promoted and those repeating Grade 1, the remainder shows the number of dropouts. Dividing this number by the total enrollment in Grade 1, 1982 will give the dropout rate:

50225 - (32007 + 16114) = 2104 / 50225 = .0419 (d)

As was seen earlier, adding the promotion, repetition, and drop out rate for a given year results in a total of 1 since all students must be in one of the three categories:

.6373	+	.3208	+	.0419	=	1.000
(p)		(r)		(d)		

As an exercise, use the procedure just described to calculate p, r, and d for Grade 2, 1983 on Table 4.4. Check your answers first by seeing if they add to 1.000. The

By now it should be clear that the models and methods for enrollment forecasting call for repetitive arithmetical calculations. These calculations can obviously be done by hand. However, what would be a tedious task for the planner is an ideal task for a computer. A spreadsheet package (such as LOTUS 1-2-3)* is very convenient for enrollment forecasting because the basic spreadsheet form is a matrix, with rows and columns. Once a planner has "written" the basic model on to a spreadsheet (in a form the computer can understand), the calculations will be automatically completed. Clear printouts of the results will help in analysis and reporting.

The following section will describe how a computer-based model for enrollment forcasting can be developed. No instructions will be given for actually creating a spreadsheet, since such instructions will vary depending on the package. Rather, the development of the computer-based model will be explained in a theoretical sense. Planners must then adapt the basic model to the software they have at hand.

*Statistical packages such as SYSTAT can also be used for enrollment forecasting procedures.

Answers for Grade 2 (1983) : (p) = .6800; (r) = .2062; (d) = .1138.

(r) = .2062; (d) = .1138.

4.5 ENROLLMENT FORECASTING WITH COMPUTERS

To begin to design a computer model for enrollment forecasting, the basic forecasting procedure must be reduced to a mathematical form which states that

> To estimate enrollments by grades for the following year, multiply the enrollment by grades in the present year by promotion, repetition, and dropout rates and add in new enrollments coming into the grades during the year.

This statement can be more easily expressed in simple algebraic notation:

$$z_{(t+1)} = A_{(t)} \cdot z_{(t)} + a_{(t+1)}$$

Where:

$z_{(t+1)}$	=	Enrollment in the following year (t = the present year) (t + 1 = the following year)
$A_{(t)}$	=	Promotion, repetition, and dropout rates for the present year
$z_{(t)}$	=	Enrollment in the present year
$a_{(t+1)}$	=	New entrants in the following year

The first task to be done is to develop a matrix of the flow rates (the $A_{(t)}$ component of the algebraic expression above). Table 4.5 on the next page is a six row by six column matrix. Both rows and columns are grades in a six-grade school. Assuming that the planner has already calculated the flow rates, the following matrix indicates where these figures are placed in the model. Then, once the matrix is complete, it can be multiplied as a whole times the $z_{(t)}$ column (present enrollment).

Table 4.5

Matrix of Transition Rates A(t)

In Symbols

Grade/Grade	1	2	3	4	5	6
1	r11	0	0	0	0	0
2	p12	r22	0	0	0	0
3	0	p23	r33	0	0	0
4	0	0	p34	r44	0	0
5	0	0	0	p45	r55	0
6	0	0	0	0	p56	r66
						(p67)
	d1	d2	d3	d4	d5	d6
p+r+d =	1	1	1	1	1	1

Here is how to interpret Table 4.5:

1. Repetition rates (the r's) are written down the main diagonal, in cells r11 to r66. The reason for this is that repetition means enrollments stay in the same grade in the next year. Thus, r11 simply means that the repetition rate for Grade 1 Year 1 will be placed the first row, first column.

2. Promotion rates (the p's) are below the r's in the diagonal. The reason for this is that when the p is multiplied by the enrollment it moves it down one row in the next year, which means ahead one grade. Thus, p12 stands for the proportion of the Grade 1 students in Year 1 that were promoted into Grade 2.

3. Drop out rates are listed along the bottom row of each column.

The symbols on Table 4.5 show where to put in the appropriate flow rates. Table 4.6 on the next page shows the same matrix with the data inserted.

the same matrix with the data inserted.

Table 4.6

Matrix of Transition Rates* A(t)

In Numbers

Grade/Grade	1	2	3	4	5	6
1	.32	0	0	0	0	0
2	.64	.21	0	0	0	0
3	0	.68	.20	0	0	0
4	0	0	.67	.14	0	0
5	0	0	0	.72	.08	0
6	0	0	0	0	.82	.03
						(.94)
	.04	.11	.14	.14	.10	.03
p+r+d =	1	1	1	1	1	1

*Rates rounded to two decimals.

Instead of the symbol r11 on Table 4.5, the actual first grade repetition rate that was computed on page 14 has been inserted on Table 4.6. (The r of .3208 was rounded off to .32.) Instead of the symbol p12, the actual promotion rate computed for Grade 1 in Year 1 of the forecast on page 14 has been inserted: .6373 (rounded to .64).

Compare Table 4.5 and Table 4.6 carefully. Make sure that you can associate each symbol with a computed value. Also recall that you can check whether the computation in Table 4.6 is accurate. Do you remember how?

(Hint: P + r + d must equal 1 for any given grade in any given year. Try adding the numbers in Table 3.6 to see if they check. If they do not, analysts recompute them.)

Once the flow rate matrix is developed, the planner is ready to create the entire model. To do this, it will be good to review again the basic mathematical expression, except the terms will be rearranged:

$$A_{(t)} \cdot z_{(t)} + a_{(t+1)} = z_{(t+1)}$$

The basic flow model shown in Figure 4.1 below will use this expression as the basis for computer-based enrollment forecasting. Notice that the $A_{(t)}$ matrix of the the model uses the same format as Tables 4.5 and 4.6.

Figure 4.1 may appear difficult to understand at first glance, but it is actually only a more detailed form of the basic mathematical expression. It shows that

The enrollments in the next year
for grades 1 - 6 [$z_{(t+1)}$],

will equal the enrollments in the current year
for grades 1 - 6 [$z_{(t)}$],

premultiplied by the matrix of flow rates (p, r, d)
for the current year [$A_{(t)}$],

and with new entrants for the following year
added in at whatever grade they occur [$a_{(t+1)}$].

added in at whatever grade they occur [$a_{(t+1)}$].

It is important to understand all of the symbols on Figure 4.1. To do that, recall that:

r, p, d	=	flow rates of repetition, promotion, and dropouts
t	=	the current year
t + 1	=	the following year
1 - 6	=	Grades 1 - 6

Thus, the symbols are interpreted like these examples:

$r_{(1t)}$	=	repetition rate for Grade 1 for the current year
$p_{(2t)}$	=	promotion rate for Grade 2 for the current year
$d_{(6t)}$	=	dropout rate for Grade 6 for the current year.
$z_{(1t)}$	=	the current enrollment for Grade 1
$a_{(2,t+1)}$	=	new entrants for Grade 2 in the following year

The creation of the computer model now calls for inputting the needed formulas to create the flow rate matrix $A_{(t)}$ (as shown on Table 4.6) and the columns for present enrollment $z_{(t)}$, new entrants $a_{(t+1)}$, and the enrollment estimates for the next year $z_{(t+1)}$. This is often called creating a <u>template</u>, or a pattern of calculations needed to complete a particular procedure. Once the template is developed, actual data can be input into the appropriate cells. The computer will then systematically carry out the calculations called for in the model. In the case of a template for enrollment forcasting, the computer will begin the calculations by multiplying three things:

1. The promotion rate by the enrollment in that level, in order to move the enrollment in that grade and that year down to the next grade (one level down) in the next year.

2. The repetition rate by the enrollment in the same level to keep a fraction of the enrollment in the same grade in the next year.

3. The dropout rate by the enrollment to take a fraction of the enrollment out of the system entirely.

Once the enrollments have been moved along to the next grade in the next year, new entrants are added into the system. (In most systems, new entrants are brought in only in the first year.) When this is completed, the computer will print out the results (estimates of the next year's enrollments) in the $z_{(t + 1)}$ column.

The final section of this chapter present an example of enrollment forecasting using actual data from a country. The example will show how the template shown in Figure 4.1 looks when actual data are inserted.

4.6 ENROLLMENT FORECASTING BY COMPUTER: AN EXAMPLE

Figure 4.2 shows the basic computer model that now has actual enrollment figures and flow rates plugged into the matrix. The example traces the flow of student enrollments from the base year 1985/86 to the following year 1986/87 for the first six primary grades.

Recall that $A_{(t)}$ is the matrix of repetition, promotion, and dropout rates for the six grades of elementary schooling in this country. The repetition rates are .206, .150, .160, etc.; they are listed down the main diagonal. The promotion rates are .735, .840,.830, etc.; these rates are listed to the left of (or underneath) the main diagonal. The dropout rates are along the bottom row, .06, .01, etc.

For simplicity, this example will exclude two things from consideration: (1) The dropout rates in the bottom row and (2) the final column of zeros in the matrix $A_{(t)}$. Having done this, the example deals only with the six rows and six columns within the inner lines of $A_{(t)}$. In other words, $A_{(t)}$ is a six-row and six-column matrix just as shown on Table 4.6.

[Important Note: On most spreadsheet programs, the matrix must be set up exactly as on Figure 4.2 with no blank rows or spaces between the rows and columns of figures. In other words, the matrix must be exactly six rows by six columns. Without this arrangement, many spreadsheet programs will not work properly.]

The numbers shown in $z_{(t)}$ are the enrollments in grades 1 to 6 in the year 1985/86. The numbers shown in $a_{(t+1)}$ are the entering enrollments to Grade 1 in 1985/86. (Note that this arrangement shows new enrollments entering only at Grade 1, but in some situations, new entrants in 1985-86 could come in at any grade).

Once the planner has inserted the correct formulas into the spreadsheet software for the enrollment flow model, the computer will perform the calculations for column $z_{(t+1)}$ using the now familiar general rule:

> Multiply each entry in the Matrix $A_{(t)}$ row by each entry in the $z_{(t)}$ column corresponding, and add in the $a_{(t+1)}$ number.

Examples:

Grade 1 = (.206)(84206) + 58551 = 75897

Grade 2 = (.735)(84206) + (.150)(66433) = 71856 + 0 = 71856

Before concluding this chapter on enrollment forecasting, a note of caution about the procedure is presented below.

Enrollment Forecasting: A Caution

Watching a spreadsheet rapidly displaying the enrollment estimates for each grade is quite enjoyable. However, it is good to keep in mind when using computer

behind the smooth analysis and forecast of numbers, there usually are assumptions (informed and systematic guesses) about how numbers of rates will work. These assumptions are not without sense or basis in reality. Usually, they result from past experience with data in the system. But they are assumptions nonetheless.

Although the enrollment flow model is widely used and generally accepted to be true, what are the assumptions that underly the model?

Assumptions

- The flow rates are stable; they do not change change over time--at least over the period of the forecast.

- The flow rates can be set or changed as a result of policy and program actions. These assumed or target rates then become stable enough to project enrollments.

In actual fact, of course, flow rates may be thought of either as historically determined or as targets which are set as statements of intent. In most developing countries, flow rates are not stable, but change over time. This does not negate the usefulness of the model. However, it does put a time limit on the validity of any projections from the flow model. As has been mentioned before, in planning, because we deal with the future, we are always dealing with probabilities, never with certainties.

APPENDIX A

USING AGE-GRADE DATA TO COMPUTE TRANSITION RATES

Sometimes planners have enrollments arranged by both age and grade, as well as numbers of repeaters (or promoteds, or dropouts). The age grade data can be used to improve estimates of flow rates. There are various models and methods for using age data to improve estimates of flow rates. A method developed by Schiefelbein (Davis, 1980) will be briefly described. This method is also used in the RTI System for Tracking Educational Progress (STEP) package.

The method requires enrollments by age and grade for two successive years. For longer forecasts where population must be forecast in order to forecast future new entrants into first grade--the method also requires (1) Population by single year age groups and (2) The age of entry into lowest grade of school system.

The method also requires some assumptions (guesses) about flow rates:

1. The process starts with guessing the repetition rate of the last year for each age group (starting with the oldest) with enrollment in the system at the highest grade.

2. As the process works back from highest to lowest grades, the number of those not promoted (for each age and grade) must be split into dropouts and repeaters. This is done on the basis of data, experience, and guessing. Sometimes if the highest age group is an aggregate (for example age 18 and above), the data may have to be broken into single year age data.

Once all the data are available and there is a basis for the assumptions and guesses, the method is mechanical to put into computer form and run. Table 4.7 illustrates the steps for the highest grades for the age groups 12 in the first year and 13 in the second year. This is based on age grade data, arranged as in Table 4.7 on the next page.

Table 4.7

Numerical Example Age/Grade Transitions

	12 year old Enrollment 1977	13 year old Enrollment 1978
Grade 6	12189	12025
Grade 5	13075	13488
Grade 4	15606	12046

Continue on to Grades 3, 2, and 1

Step 1: Start at grade 6 and assume a repetition rate of .062. Multiply 12189 X .062 = 756 repeaters.

Step 2: Subtract this 756 from the enrollment in the next age grade: 12025 - 756 = 11269 promoted from the 13075 enrolled in fifth grade in prior year.

Step 3: Subtract 11269 from 13075 enrolled in Grade 5: 13075 - 11269 = 1806 not promoted from Grade 5.

Step 4: Split Not Promoted 1806 into Repeaters and Dropouts. In this example assume .98 (1806) = 1770 Repeaters; .02(1806) = 30 Dropouts.

Step 5: Subtract 1770 repeaters from the 1978 Grade 5 Enrollment: 13488 - 1770 = 11718, which is number promoted from Grade 4 in prior year.

Step 6: Add 11718 (promoted) and 1770 (repeated) and subtract from 15606 to get number not promoted from Grade 4: 11718 + 1770 = 13488 - 15606 = 2118 not promoted from Grade 4.

Step 7 Divide non-promoted in Repeaters and dropouts as before. Continue on to grades 3, 2, and 1.

APPENDIX B

COHORT ANALYSIS

Figure 4.3 on the next page shows an illustration of cohort analysis. The record of students' promotion, repetition, and dropout rates was calculated based on the reports of parents as to whether the student was promoted, repeated, or dropped out in any given grade/year transition. The history can then be compiled to give a simulated cohort history of p's, r's, and d's. These estimates can then be used in the basic flow matrix.

However, there are difficulties with cohort history transitions when they are applied in some models. Because the history is carved out of the middle of a longer time segment, it may be very atypical and give distorted estimates.

Some models use the estimated cohort transitions to compute a variety of output indicators:

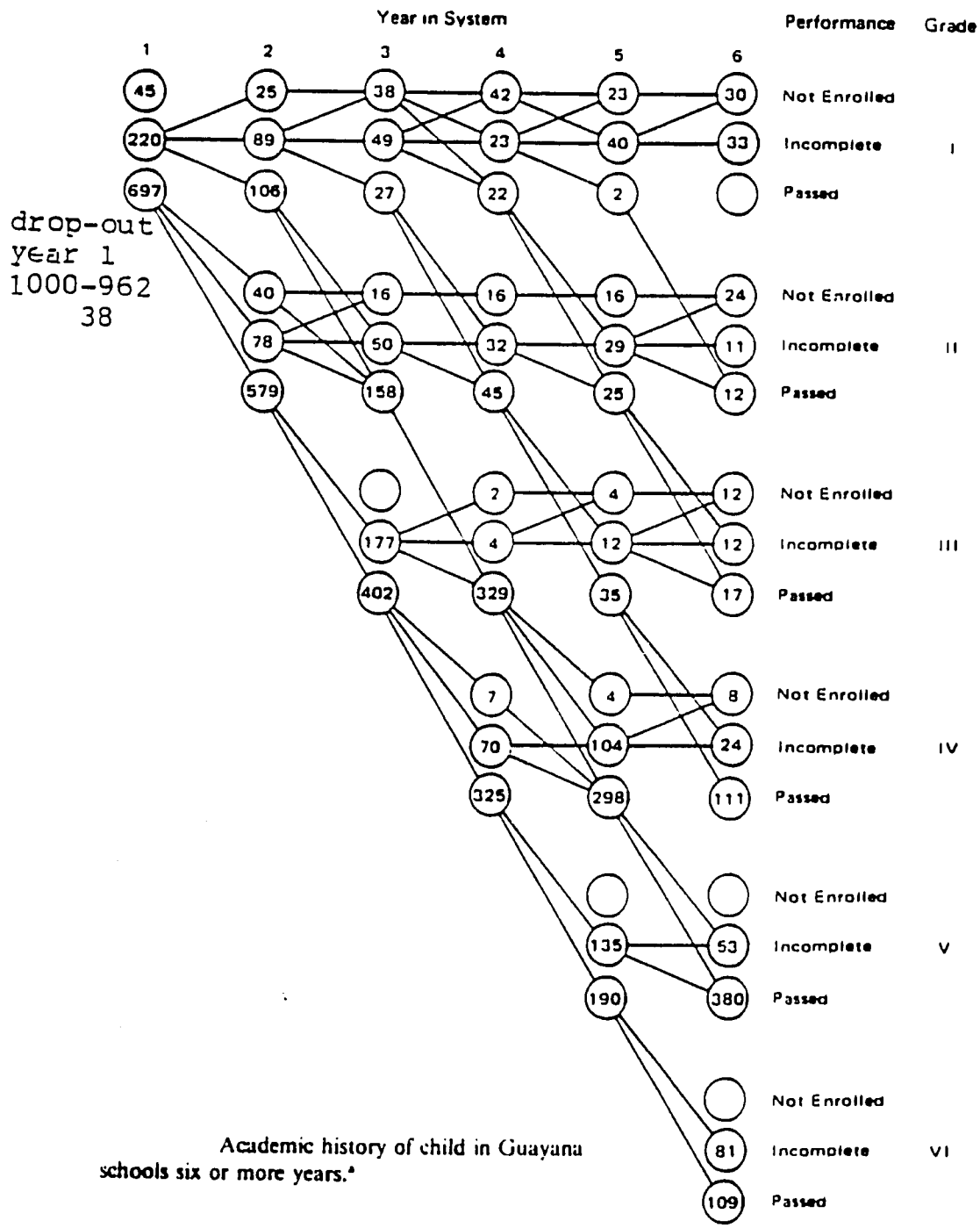
1. Number who eventually graduate: In the example, this cannot be computed unless all children in the cohort are followed to their final school status. This might take 12 or even 15 years in a school system with high wastage and overage children. If those who "passed" in the last three grades on Figure 3.3 all eventually graduated, the number would be 600, and the rate would be 600/1000 = .6.

2. Number who graduate in normal number of years (6 years in a six grade primary school). Based upon the data on Figure 4.3, this number would be 109/1000 = .11.

3. Student years per graduate, or the total years spent by all students divided by the number of graduates. These figures (like all "cohort" flow rates) can be very unstable over a short time span. For example, assume that a total of 8000 years of education (including dropouts and repeaters) was needed to produce the 600 graduates. The Student Years per graduate would be 8000 / 600 = 13 years, which means twice the normal 6 years of the level.

Figure 4.3

Cohort Simulated Flow in Six Grade System



Academic history of child in Guayana schools six or more years.[a]

Chapter 5

MANPOWER PLANNING

5.0 INTRODUCTION

The economic forecasts that directly concern education planners in their broader role of human resource development planners are:

A. Forecasts of economic activity and growth in product, value added, or income. This covers the sum total of the goods and services produced by the national economy as measured in GNP (Gross National Product), GDP (Gross Domestic Product), and National Income. The total product is usually forecast for each of the main sectors of the economy.

B. Forecasts of EAP (Economically Active Population), labor force or work force; forecasts of employment, classified by sectors of the economy; and manpower demand forecasts, by occupation and education levels.

C. Productivity, or product per head, derived from product and employment. These forecasts are guided by estimated changes in technology and capital investment, and changes in the way enterprises organize for production.

D. Other forecasts detailed by industries or occupations of special interest, or by groups in the population; by earnings; or by income levels.

Educational planners use economic forecasts primarily in setting goals for the education/training system in terms of product changes in the economy; for setting targets in terms of manpower demand, and requirements for educated and trained manpower estimated on the basis of demand and forecast supply (output) of education/training systems. The economic forecasts are usually accomplished by economic analysts in the national planning office, rather than planners in a ministry of education, but educational planners must be familiar enough with the economic forecasts to evaluate their general credibility and to incorporate them into human resource development forecasts and plans.

This chapter introduces the topic of macro-level manpower planning and outlines a basic six-step forecasting method. Included are descriptions of industry and

occupational classification systems and a discussion of alternative models. Data from manpower planning case studies of Chile and The Dominican Republic (Davis, 1980) are used to illustrate the material presented in this chapter.

5.1 MANPOWER PLANNING APPROACHES AND THE BASIC FORECASTING METHODS

Figure 5.1 on the next page outlines the simple models, methods, and basic information required to forecast manpower requirements and the output, or supply, of education/ training systems. The term Manpower Demand refers to the forecasts of manpower requirements, and Educational Supply refers to the forecasts of enrollment flows and graduates of schools and training institutions. Demand is not a true labor market demand estimate, but simply projected expectations of labor force growth derived from projected growth of product.

When the methodologies shown on Figure 5.1 are applied in actual planning, there are slight variations as steps are combined, or the ordering of steps may be different. Models and methods are never slavishly followed. The work depends on the information available, the pressure which deadlines impose on the planner, the competence and working style of analysts, and the level of precision required in the results.

Figure 5.1 outlines other information which may be available to the planner, information that will permit a deeper and enriched analysis, and an improved result. The basic method for forecasting manpower requirements is primarily useful in providing a structure into which much more detailed and precise information can be incorporated, and results refined in subsequent iterations.

The section that follows will run through the steps required to project manpower requirements using the basic method. The object of the projection is to provide an estimate of the requirements for educated and trained manpower during the future years of a plan. The projected manpower requirements can be expressed in terms of education and training requirements, which can then be compared to projections of supply from the education/training system. The resulting deficits or surpluses can then be used to guide the planning of education and training.

Figure 5.1

Outline for Manpower Planning

I. AGGREGATE LEVEL FORECASTS

1. BASIC METHOD

Steps

a) Product forecast by sectors (plan targets)
$P \longrightarrow$

b) Productivity forecasts
$P_{p.w.}$ = product per worker

c) $\frac{a}{b} = E$: Employment by sectors

d) E distributed sectors by occupations

e) Occupation distributed by education (levels and programs).

f) Education "demand" aggregated

2. Alternate Method (Growth in Occupation)

X_{ij} = number in occupation (i); sector (j) in the base year

D_{ij} = future requirements in occupation (i), sector (j)

Steps

a) Project D_{ij}

$$D_{it} = \sum_{j=1}^{n} (X_{ij})_t \; e^{\left[(b_{ij}\, r_{pj} + Y_{Lj})\right]^t}$$

b_{ij} = rate of change of workers in occupation (i) as productivity in sector (j) changes

r_{pj} = National data historical trend productivity rate

Y_{Lj} = Growth in numbers of workers trend

e = Form of the growth curve is exponential

b) Distribute by Education levels and Programs

c) Aggregate Education "Demand"

3. Apply aggregate level ratios as checks, e.g.,

a) Participation rates
b) Demographic rates

II. DISAGGREGATED (PIECEMEAL) REQUIREMENTS (TO SUPPLEMENT AGGREGATE FORECASTS)

1. Service Sector/Government
   A. Population based service norms/ratios
      a) Education (link to supply)
         i) Social Policy (coverage, demographic)
         ii) Education policy Program staffing, e.g. t/p ratios
      b) Health Services
         i) Coverage needs, demand
         ii) Delivery systems/norms/staffing i.e. Beds/Doctors/Nurses/Parameds
      c) Other Government Services
         i) Defense (numbers, organization manning tables)
         ii) Police, fire sanitation (state, municipal)

2. Core Industry Arrays
   Input/Output, Forward/Backward Linkages
   a) Scale/Technology/manning tables (present and future)

   b) Establishment surveys Employment (Present & Future Estimates Wages and Salaries)
      X occupation, given market share,
      prices, scale, technology

3. Small Industries and Pvt. Services
   Linked to other industries & service needs
   Linked to population served ratios & technology
   Linked to income & effective market

4. Agriculture
   Crops/Acreage
   Land Tenure Patterns
   Cultivation Patterns
   Export, World Demand, Export Policies,
   Prices Domestic, Numbers, Diet, Income

5. Fill details in cells and check with Aggregates from I. Final iteration deficit Phase I.

   Demand: Base stock + wastage - supply = Deficit
   Interaction in subsequent iteration phases

III. DEMAND/SUPPLY EFFECTS ON FORECASTS

1. General Government Goals/Policies Analysis
   A Growth Rates economy
      a) investment, monetary fiscal
      b) Employment plans policies
   B. National Education: Demographic/Social targets
      a) access
      b) flow
      c) output, programs/issues
   C. Education Sector Policies
      a) input norms
      b) costs
      c) resource constraints revenue policies/finance H.R. lags(e.g.trained staff)
      d) Admissions/scholarship policies, subventions, systems & institutions
         i) influence access & flow
         ii) Influence incentives choices

2. Demand/Supply/Price Interactions
   Econ.
   A. Labor Market Information
      a) job opportunity
         i) openings, promotion
         ii) earnings (wages, salaries)*
         iii) other returns (phychic)
      b) Guidance Information
         Voc/Career choice
         Educ/Career choice
   B. Rate of Return Studies*
      a) Earnings profiles
      b) Employment probability

3. Social/Cultural Values
   a) National
   b) Community
   c) Family
   d) Individual

Value of Education

*Effects on Demand & Supply Choices (Elasticities if possible)

5.2 THE BASIC PROJECTION METHODOLOGY: OVERVIEW

The base of a manpower demand/supply forecast is the work force in the base year, classified according to sector of the economy and occupation. From this point, the basic method proceeds in a certain sequence (although the sequence may be modified when applied):

Starting Point: Base Year Work Force by Sectors and Occupations

Step One: Project Employment to the Target Year

Employment is projected to the target year by sectors. This step has three parts:

a. Product is projected by sectors of the economy.
b. Product per worker, or productivity, is projected.
c. Product is divided by productivity to give the number of workers, or employment.

Step Two: Distribute Employment by Occupations Within Sectors

Employment, projected in step one, is distributed within sectors, by Occupations.

Step Three: Estimate Education Requirements of the Work Force

Within occupations, the education or training levels of workers is set. We have the future year or plan year work force, described by sectors, occupations, and educational levels.

(The first three steps are called projecting the Demand. This means projecting the future educational needs of the work force.)

Step Four: Forecast Supply

We now project the output or supply of all education and training institutions (and allow some additional supply from those trained on-the-job).

Step Five: Determine Surplus/Deficit of Trained Workers

We can now subtract the supply of educated and trained people from the projected demand (Step 3). This gives the deficit or surplus of educated or trained people

that are necessary to attain the projected growth of product and productivity in the economy. In theory the education and training system can be adjusted to produce these trained people, according to the deficit or surplus.

Step Six: Account for Wastage

There is one other small step. During the course of the plan, workers in the original work force and workers graduated and employed in the work force during the plan period, change jobs, retire or die. This must be taken into account. It is called wastage. This is done by applying work force wastage rates to the original work force and to the graduates who come into the work force in the plan period. (These rates are based on death rates and retirement rates.) Once the wastage or outflow is estimated, there are two ways of getting the final deficit or surplus:

a. Add the wastage to the final demand and then subtract the base year work force and the supply:

Deficit = Demand + Wastage - Base Year Stock - Supply

b. Subtract wastage from Base year. Then subtract wastage from supply and subtract the result from Demand to get Deficit:

Deficit = Demand - [Base Year - Wastage] - [Supply- Wastage]

(Note that the two formulas above are algebraically equivalent.)

The final step then yields surplus or deficits with wastage accounted for.

Each of the steps will be described in detail in the next section. It should be noted that although the steps are straight-forward--almost mechanical--the assumptions are by no means mechanical. Assumptions are made about changes in product, productivity, occupational distributions within sectors, educational distributions within occupations, and in the setting of wastage rates for future years. Each of these could be subject to error or challenge because there is no automatic way to read the future. Later sections will examine some sources for guiding these assumptions.

5.3 THE BASIC PROJECTION: DESCRIPTION

Starting Point

Base Year Work Force by Sectors and Occupations

The first requirement is the base-year work force by economic sector (Agriculture, Industry, etc.), occupation (Manager, Engineer, etc.), and education level. Sector and occupation classification are discussed in the final sections of this unit.

Figure 5.2 below shows the format for the distribution of the work force by occupation (in the rows of the table), and sectors of the economy in the columns of the table. Figure 5.2 has the work force distributed for the base year, x, and the target year, n. At this point only the entries in column x, for the base year will be filled in.

Figure 5.3 on the next page shows the format for the distribution of the work force by occupation and educational attainment for the base year. We now have the basic information required for starting the manpower projection. Note that the occupational classification is the common link between the sector classification (Figure 5.2) and the educational classification (Figure 5.3).

Figure 5.2

Distribution of Work Force by Occupation and Sector (Base Year x and Target Year n) of Economic Activity

Sectors	Total No		Total %		Agriculture[a]		Mining		Manufacturing		Services	
Occupation	x	n	x	n	x	n	x	n	x	n	x	n
0 Professional												
1 Administrative												
2 Clerical												
3 Sales												
5 Farmers												
7 Artisan-Craftsmen												
8 Operators												
9 Service Workers												

[a]Each sector column inlcudes number and percent of workers in occupations within the sector

Figure 5.3

Educational Attainment by Occupation in Base Year x

Occupation	Col. 1 Total	Col. 2 University graduate studies complete (19 yrs.)	Col 3 First degree level univer-sity (17 yrs.) Professional and higher normal	Col. 4 Sub-pro-fessional (15 yrs.)	Col. 5 Tech. complete (12 yrs.)	Col. 6 Secondary general	Col. 7 Lower normal complete (12 yrs.)	Col. 8 Primary complete (6)	Col. 9 Literacy grade (4 - 6
0. Professional/science No. %						Includes those who fin-ish 12th grade only plus those who finish 12th grade and go on to col 2 3, and 4 groups		Includes those who finish primary plus those who go on to cols 5, 6 and 7 groups	Includes those who finish at least 4 grades plus col 8 group
1. Administrative/managers No. %									
2. Clerical No. %									
7. Artisans No. %									
9. Operators No. %									

Note: Educational levels may be grouped in various ways. The basis is arbitrary. Here the groups are
Col. 2 includes all who finish through 19th grade.
Col. 3 includes all who finish 17th through 19th grade.
Col. 4 includes all who finish 15th grade sub-professional.
Col. 5 includes all who finish only 12th grade technical training plus the proportion of those who finish this track and attend higher education (both sub-professional and professional, i.e. lines 3 and 4).
Col. 6 includes all those who finish only 12th grade general preparation plus those who go on to higher education from this track (because schools must produce graduates who terminate at 12 and graduates who continue on.)
Col. 7 includes all those who finish only 12th grade lower normal plus that proportion from this track who go on to higher education.
Col. 8 includes all those who finish primary, plus those who finish grades 7 up to 12 (in the secondary tracks) plus the totals of 5, 6, and 7.
Col. 9 includes 8, plus those who finish at least 4 grades but less than a complete 6 years of primary.

National projections are mainly useful for planning in national, centralized education and training systems, and where planning is organized under a national planning office. This is the situation in most countries in Europe, Africa and Latin America. In the U.S., most manpower forecasts and plans are made for subnational levels, regional, state, or metropolitan area labor markets.

The prime source of information is the national census, which collects information on the work force and tables it by sectors, occupations, and sometimes by educational attainment. There may be a number of limitations on the information as commonly presented in census tables. The information may not be arrayed in the form of occupations by sectors, as in Figure 5.2, or in occupations by educational attainment levels, as in Figure 5.3. The information can be refined or further processed in special runs to meet the planner's specific needs.

A common problem is that the data may be too highly aggregated for use. Only single-digit classifications of occupation and sector (as in Figure 5.2) may be available, and special runs may be necessary to get the data into finer occupational or sectoral categories. The same may be true of educational categories.

Also, census information may not be in the appropriate format; different cross-tabulations may be required, or the information may need to be updated. The first plan year may be 1985 and the most recent census 1980. In this case, the census must be brought forward five years to current date, or to the base year of the planning exercise. The planner may use survey data from special work-force studies, e.g., skill surveys or establishment surveys. There may also be detailed occupational information for a specific large industry or enterprise. The basic census data can be updated, supplemented, and made more detailed with results from other studies and surveys of the work force.

Step One

Project Employment to the Target Year

a. Product has been projected from base year to target year by sector in overall development plan.

b. $\frac{\text{Product}}{\text{Workers}}$ = productivity, derived from analysis of trend years. On the basis of past trends or future expectations of increase or decrease in sectors, productivity rates are projected from base to target year.

c. Employment is then projected by sectors for target years, by division of a/b.

The projections of product and productivity may be taken directly from development plan estimates. Or the projections may be based on past or historical trends of product growth extrapolated to future years. Or the projection may rest on a combination of the two things, with trends modified by information on plant capacity, technology, and investment plans for expansion. Comparative experience, the growth experienced in other countries or economies, may also be used to guide the projections.

The simplicity of the first step is shown in the following symbols. The aim is to project employment from year x to year n. Employment, symbolized E, equals P (product) divided by P p.w. (product per worker).

$$nEx = \frac{P \quad = \text{(Product (Projected))}}{P_{p.w.} \quad = \text{(Product Per Worker)}} = \text{Projected employment } x...n$$

Product and product per worker are projected for each year x to n. Employment is projected separately each year for each sector. Sector product P, and product per worker $P_{p.w.}$ may be taken directly from the national development plan or estimated on the basis of performance of the economy, or assumed on the basis of informed judgments about development potential in sectors or specific industries. The growth experienced in other similar situations in other countries, or in other regions or industries in the same country, may also be used to guide the projections.

From the outset the results may depend on how well the national development planners have estimated product in sectors for the future year. This is a crucial estimate,

and it underlies the employment projection, as it underlies almost all other attempts to derive manpower demand from the economy. The adequacy of any employment projection depends on the adequacy of the basic projection of growth in product.

Base-year estimates of product and productivity can be in error. If product and productivity are projected on the basis of a trend extrapolated from past years, the adequacy or representativeness of the base years is in question. If the product/productivity ratio is projected on the basis of what has happened in the sectors and assumptions about probable changes, the adequacy of these assumptions must be examined. The units in which the product is expressed can also be troublesome. Monetary values may have to be adjusted to a base year, and when this is difficult and no deflator or price indexes are available, the product may have to be expressed in units of physical output.

Census data usually furnish the basic numbers in the work-force employed and unemployed. The census data may have to be updated to the base year by an establishment survey which covers new employment. The establishment survey may also be designed to get management's best estimate of likely expansion and future employment. Hence, it may furnish one basis of work-force estimates for future years.

When countries or industries are growing and changing, the establishment survey will have shortcomings as a basis for future year estimates. The survey covers only existing industries. Industries that may not yet exist or be identified as possibilities early in a plan period may come into existence and grow rapidly during the plan period.

In some areas, planning may have reached the stage where most major industrial possibilities have been studied. Employment for these future industries may have been estimated through feasibility studies which also cover sources of raw material, investment and financing, technology, production problems, and market and sales forecasts. On the basis of these studies, employment can be directly estimated for future years.

Other Methods for Estimating Employment:

Planners may use a general sectoral employment estimate as a first approximation, but usually they go beyond this and work on estimates for individual industries or groups of

industries within the sector. A general ratio of product to product-per-worker may be used to derive a first estimate of employment for an entire heavy industry group. Then, detailed employment estimates may be prepared for each individual industry, and the results adjusted and compared with the totals.

Product or productivity may be derived from investment ratios, or estimated on the basis of market demand, minimal effective size of the plant, and assumptions about the percentage of capacity at which it will operate. Product for one industry may be derived from another, e.g., oxygen from steel. Detailed input-output relationships may be built up to derive product for one industry, group of industries, or sector, from another.

One sectoral grouping may be major industries, a second group made up of industries supplying goods and services to the major industries, and a third group made up of industries supplying goods and services to households. The product and employment of the group of industries that supply goods and services to the major industries could be related to growth in the major industries. The product and employment of the industries supplying goods and services to households could be related to growth in population, households, household income, and expenditure.

For light industries, and in the service and commerce sectors, employment may be derived from a ratio to the total population projected for the area. From the size of the population, household income and the effective market, an estimate of the required output can be made; and from output, the number of workers required. Industry fact sheets or Manning Tables (detailed estimates of number and kinds of workers by jobs) can provide employment estimates and occupational distributions according to scale and technology of the industry. This detailed information may be summed into the general employment estimates for target years.

The result of the first step is employment in the target year by sector of economic activity.

Input/Output Tables and Models:

Input/output tables which show the input of one industry to another industry and to final demand can be converted to coefficients and solved as a system of linear equations to yield the output of the industries required to meet some projected final demand target.

The significant point is that there are many alternative ways to derive employment by industry sectors; and employment by sectors is the result of the first step of the basic method. The first step yields employment for each sector. Input/output tables can be used to estimate product, productivity, and employment.

Methods for Checking Employment Estimates:

When the data are put into the tables, several check ratios can be applied. Employment can be expanded to population economically active by estimating unemployment and adding it in. The economically active population as a percentage of the total population can serve as a control check. This participation rate can be further broken down for age groups, education levels, and sex, and compared with experience in the past and with data from other similar countries. The percentage distribution of employment by sectors can also serve as a check, e.g., the percentage in services as compared to the percentage in industry. Gross overestimates and underestimates will show up. The percentage of the work force found in broad sectors of the economy--for example, in the primary sector of agriculture--is also a check.

Supplementary Reading on Sectoral and Occupational Classification Schemes:

Systematic tabulating of sectoral and occupational data is required to work with manpower data. The sectoral and occupational data are tabulated according to standard systems for classifying sectors and occupations. One-digit sectoral and occupational classification schemes were illustrated in Figure 5.2. In the Chilean case, one-digit classifications will be used, although the sectors and occupations are slightly different. In manpower analysis the planner may attempt to work with more refined classification schemes that use two or three digits and break industries and occupations into more detailed groupings. The classification of the work force by sectors and occupations is central to aggregate-level manpower forecasting and requires some additional explanation.

1. Sectoral or Industry Classification

The United Nations (UN, ST/STAT/Ser.,/27) defines what it refers to as industry (which is also called branch of economic activity or sector when a one-digit, large group is referred to) as follows:

> Industry refers to the kind of establishment in which the person works (or worked if unemployed). For purposes of international comparability, it is recommended that countries adopt the International Standard Industrial Classification of All Economic Activities (ISIC) most recently approved by the United Nations, or that they tabulate their statistical data so that the categories can be regrouped in accordance with the Standard Classification or at least with the divisions (one digit) of this classification.

The ISIC (United Nations Statistical Office) provides a taxonomy for classifying economically active civil populations. The major divisions or sectors are: (1) Mining; (2-3) Manufacturing; (4) Construction; (5) Public Utilities; (6) Commerce; (7) Transportation/Communications; (8) Services; (9) Not Specified.

The eight divisions are divided into major groups (two-digit code) and subgroups (three-digit). The standard classification is usually modified for any given country. Government may be added as one of the main divisions. Services may be separated into more than one main division, e.g. private and government services.

The basic scheme used in manpower planning in many countries is the one developed and used in the national census and statistics office, as in the Chilean case. Other government agencies, for example, the Ministry of Industry or the Ministry of Economics, may have slightly different schemes. In the United States, the Bureau of the Budget has developed the Standard Industrial Classification, and the industrial classification system developed by the Bureau of the Census is also used. There are systematic procedures for translating from one system to another. These are called <u>crosswalks</u>.

One of the standard classification systems may be used as a model, with appropriate modifications, according to the economy of the particular country. In any given country application there are usually modifications made, but there are disadvantages when a country departs too widely from one of the standard classifications. Comparability is difficult, and retabulation may require going back to the original census schedules, a difficult and expensive task. Still, simplification of the categories may be necessary because of deficient data. A simple three-sector breakdown (industry, agriculture, and services) may be all that is possible.

2. Occupational Classification

On occupations the United Nations' Principles and Recommendations for National Population Census states:

> Occupation refers to the kind of work done by the person employed (or performed previously by the unemployed) irrespective of the branch of economic activity, or the status (as employer, employee, etc.) in which the person should be classified. For purposes of international comparisons, each country should provide for the necessary subdivisions of its occupational classifications to make possible the classifying or reclassifying of the data in conformity with the latest education of the International Standard Classification of Occupation (ISCO).

The ISCO classes economically active civil populations into nine major occupational groups. The major groups are subdivided into 70 minor groups (two-digit level) which are subdivided into 200 unit groups (three-digit level). Unit groups can be broken down further. Other systems of occupational classification are used. The U.S. Bureau of Labor Statistics has used a classification system based on the Dictionary of Occupational Titles. The Inter-American Statistical Institute, on the basis of the 1950 Census of America and ISCO, worked out a classification scheme for the 1960 Census of America (COTA-1975). This system has been developed in various ways over the years; in an actual country situation, the basic system is modified according to the prevailing occupational structure. Some country systems tend to follow the original model too blindly.

Updating Occupational Data:

Occupational classifications tend to become dated rather quickly, as new technologies and administrative hierarchies develop and new job classifications and occupational titles are created to fit the situation. Not only does the nature of the work change, but the educational requirements--and these are the two main bases for occupational classification--also change. The change is almost always upward, e.g., more training and education are required. In the United States, the Occupational Outlook Handbook is designed to keep new workers and guidance counselors informed about recent and possible changes in demand for workers, new work patterns, and new

education and training requirements. The compilers state that some of the information will be out-of-date before the edition appears in print. The Dictionary of Occupational Titles is the best reference for complete coverage and, like the Occupational Outlook Handbook, has been revised in recent editions. The most recent Occupational Outlook Handbook is the 1988-89 edition (Bureau of Labor Statistics, U.S. Department of Labor).

The Dictionary of Occupational Titles (fourth edition: 1977 and DOT supplements in 1979, 1981, and 1986) relates the classification system in the DOT to other government occupational language systems. The other occupational language systems are those of the Bureau of the Census, the Bureau of Labor Statistics, the U.S. Office of Education, and the Office of Management and Budget Standard Occupational Classification (SOC) program.

The need for constant reclassification updating and editing requires a modern data base management system. The NOICC (1984) occupational information system has a demonstration package in D-Base which can be easily updated. In the LDCs the decennial census is almost the sole source of manpower data. Often detailed sector/occupation tables do not appear until midway through the censal period. Even an advancing country like Mexico did not have detailed tables until five years after the 1980 census (Mexico, 1985). In the United States, the heart of the system is the OES I-O Matrix of industries.

Summary Comment on Step One:

The first step in the basic method of manpower forecasting is to project employment on the basis of product and product-per-worker. There was nothing special in the projection model or process, in that it followed the same basic methodologies of population projection or enrollment projection. The aspect of manpower forecasting that is distinctive is the classification of workers by occupation, sector, and educational levels. Systems for classifying workers have been discussed. A following section on references and further reading will provide additional information.

Step Two

Distribute Employment by Occupations Within Sectors

Step two of the basic method distributes the employment (projected in step one) by occupations within sectors. The base-year distribution of employment by occupations within sectors is provided by census tables supplemented by survey and specific sectoral or industry employment data. This distribution is used as a guide for distributing employment by occupations in the target year. "Distributing" simply means multiplying the number in the sectoral work force (employed and unemployed) by per cents or share coefficients for each occupation. For the years of the projection, the occupational distribution will change. The task is to analyze the factors which will affect this occupational distribution, and to make assumptions as to how it will change.

There will always be growth in occupations requiring more employment and training, but the challenge is to develop a systematic method for projecting these changes. The basic method is systematic, but it is not free of subjectivity. The changes are based on the analyst's judgment. The issue, then, as in any projection method, is how adequate are the assumptions on which these changes are based.

U.S. national level projections begin from a basis of assumptions that the basic structure of the economy, technology, and employment will not change. In the Chilean case, for example, occupational distributions were based on those found in European countries, and the assumption is that Chile will follow the European experience closely enough so that the projections will serve. This comparative assumption is widely used (Dominican Republic, 1980) and misused (Puerto Rico, 1955).

Setting Occupational Distribution:

In Step two, the percentage of workers in each occupation within each sector is set for the future year. This setting can be based on experience, opinion, and percentage distributions in the work forces of other countries. The simplest procedure is to assume that the same base-year percentages will hold in the target year. Usually, however, the planner takes into account the growth

projected for the different sectors and industries from base to target year, the expectations of changes of inputs in labor and capital, and the experience of other countries, in order to set a reasonable distribution for the target year. The planner may also take into account the change in occupational structure in past years in the country and its influence on output. The major guidelines come from expected growth in certain key sectors, as assumed or projected in the overall plan. If growth is planned for the manufacturing industry, the percentages of professionals (0), administrators (1), artisan-craftsmen (7), and operators (8) might be raised in the manufacturing sector. How much would they be raised? There is no analytic answer to this. It would depend on how high the existing percentages are for the country in the base year, and how high they are in other countries that have similar outputs or industrial structures. It is sometimes possible to select a more developed country and to move the percentages up toward this target.

Distributions of percentages in occupations have been published for various countries. OECD, Occupational and Educational Structures of the Labor Force and Levels of Economic Development (1970) is also a source of guidance.

The study by Horowitz, Zymelman, and Herrnstadt, Manpower Requirements for Planning, Statistical Tables (1966) is old, but still useful, as is Zymelman's "Productivity, Skills and Education in Manufacturing Industries" (in Planning for Advanced Skills and Technologies, United Nations, New York, 1969). (These sources also provide the basis for setting educational attainment levels by occupations, which is the next step of the basic method.)

Since all distribution tables are soon outdated and useful only as a starting point, new approaches treat manpower forecasting as an information system that must be updated constantly. For example, a spreadsheet model where both assumptions, distribution structures and input numbers can be changed and the whole worksheet recalculated and rerun to get new demand, supply and deficits as surpluses to serve as plan targets. Sensitivity analysis and goal seeking trace blocks can be used to develop and test hypotheses based on the runs and results. The process never stops. Data must be constantly gathered, regathered and the models run.

In Step two, the occupational percentages set for the target year are multiplied by the employment for a specific sector or industry projected in step one. A few cells from Figure 5.4 on the next page will suffice as an example. This part of the step is purely mechanical. The setting of the percentage distribution requires experience, judgment, and sometimes luck.

Figure 5.4

Number in a Given Sector, by Occupation, in Target Year n

Occupation	Total employment in sector, year n	Occupation %, in year n	No. in sector/occupation year n
	200,000		
0 Professionals	.	3	6,000[a]
1 Managers/Admin.	.	2	4,000
.	.	.	.
.	.	.	
.	.	.	
Total		100	

[a] .03 x 200,000 = 6,000

Occupational totals can then be summed for sectors, and the result is a table showing the numbers of workers in each occupation for the target year. In Figure 5.4 the occupational distribution for the target year is set. The occupational percentages are multiplied by employment projected for the sector in step one, and the result shows the number of workers in each occupation for the various sectors. This is shown more completely in the Chile case as described in Davis (1980).

As in any projection exercise, alternative occupational settings may be hypothesized. This will yield alternative projection series. For example, one projection might show occupational distributions unchanged from base year to target year; and a second projection may assume a large increase in occupations requiring more education and training. Under the newer information system and modeling approaches these are not just set of two or three alternatives. The computer allows us to try hundreds of forecasts.

Step Three

Estimate Educational Requirements of the Work Force

The aim in step three is to estimate the educational requirements of the work force in the target year. (See Figure 5.5 below.)

Figure 5.5
Educational Attainment by Occupation in Target Year n

Occupation	Univ. grad. studies complete (19 yrs.)		First degree level university (17 yrs.)		Sub-professional.....15 yrs.		Literacy grade 4-6		
	Hyp. I	Hyp. II	Hyp. I	Hyp. II	Hyp. I	Hyp. II	Hyp. I	Hyp. II	Totals[c]
0 Professionals (Total = 100,000)	5,000[a]	7,000[b]							
1 Administrators									
.									
.									
.									
7 Artisans									
8 Operators									
.									
.									
.									
.									
Total[d]									

[a] .05 x 100,000
[b] .07 x 100,000
[c] Hyp. I + Hyp. II = 100,000
[d] Hyp. I ≠ Hyp. II

Distribution of Education by Occupations:

The educational attainment percentages for the target year are then set for the various occupations. The general strategy may be to raise university and college attainments in the first two occupational categories, and to raise technical secondary-school training attainments for craftsmen and operators. For example, if the professional group having university-level preparation in the base year is 65 percent, the strategy might be to raise this to 90 percent for the target year. This is a substantial raise and will create a large demand on the educational system, inasmuch as the professional work force in the target year

will be a much larger base number on which to levy the increase. This is so because the total population, including the work force, will almost certainly be growing during the plan period, and the percentage of professionals in the work force will have been raised in step two.

The same sources used for guiding the setting of occupational distributions, which were cited in the previous step, can also be used for setting educational attainment levels by occupations (see OECD, Occupational and Educational Structures of the Labour Force, and United Nations, Planning for Advanced Skills and Technologies). Care must be exercised in choosing comparative models. There is no standard educational attainment unit, although years completed is most often used. One country may require much more formal education than another country for a comparable occupation.

Estimating Education-Occupation Distribution:

Alternate education-occupational settings may be used for the target year. In one set, the educational attainment coefficients of the base year may be maintained into the target year. Sets of educational attainment coefficients can be applied to yield different sets of alternative projections. Before coefficients are raised the consequences must be checked against the capacity and likely revenues available for expanding the educational and training facilities. Demand and supply projections are not independent. One can be used to check the other. The final operation in step three would appear as in Figure 5.6.

Step three yields the target-year estimate of educated people required in the work force. This comes from summing each educational level down the list of occupations. The result of this step is the so called demand for educated and trained people in the work force. It is clear that this is not so much a "demand" as it is a target or a goal set by planners.

Alternative Steps in Applying the Basic Manpower Requirements:

Beyond Step three, there are alternative ways of arriving at a final demand and supply estimate for educated manpower in the work force, and the surplus or deficits in trained manpower.

Step Four

Forecast Supply

This step calls for forecasting the supply: the output of graduates or trained workers from the education and training system. These forecasts are based on the enrollment forecasts covered in Chapter Four.

Step Five

Determine Surplus/Deficit of Trained Workers

The forecasted supply in step four is then subtracted from the forecasted demand in step three, and the result is the balance (the surplus or deficit) of educated and trained workers. Note however that this has not taken into account the fact that over the period of the plan or forecast, workers retire, die or change jobs, and even new graduates entering the work force may leave it. So the outflow from the original stock of manpower and the reduction in the inflow of supply must be calculated and put into the balance to calculate the final deficit. There are two alternative ways to do this:

Alternative One

Formula:

Deficit =
Demand - [Stock-Wastage] - [Supply - Wastage]

a. Reduce the base-year stock of educated workers by applying wastage rates. Wastage is loss from death and retirement over the years of the forecast.

b. Reduce the supply coming in to the work force by Supply Wastage.

c. Subtract a. and b. from Demand to get deficit. This is essentially the procedure followed in a Dominican manpower case (see Figure 5.6 on the next page).

Figure 5.6

Application of Outflow Rates to an Age-Stratified Occupational Group (Base Year x to Target Year n)

Occupation: 0 Professional

Age group	Stock,[a] x	Outflow rate,[b] x to x + 5	Outflow x to x + 5	Stock, x + 5	Stock, n
30-34	3,000	0.0282	85[c]	2915[d]	
35-39	5,000	0.0375	188	4812	
40-44	6,000	0.0625			
45-49	7,000	0.0915			
.					
.					
.					
70-74	500	0.5081			
75+	120	0.6429			
Total outflows			w_1[d] Period 1, Total	w_2 Period 2, Total	

Note

Repeat process from x + 5 to x + 10, x + 10 to x + 15, until year n is reached.

[a] fictitious

[b] from work force life table

[c] 0.0282 x 3000

[d] 3000 - 85

Note. The outflow by occupation can be converted to wastage by educational attainment levels because the educational attainment for the occupation is known in year x.

Alternative Two

Formula:

Deficit or Surplus =
[Demand + Stock Wastage + Supply Wastage] - [Stock + Supply]

Alternatively, the planner can project the demand, calculate stock and supply wastage and add it to demand, and then subtract stock and projected supply. The usual order is to project supply and then estimate supply wastage, based on the age profiles of the newly graduating workers.

Step Six

Account for Wastage

In the step five alternatives above, outflows (workers leaving the work force) are estimated and added to the demand requirement in the target year projection. Outflows from stock are estimated by applying wastage rates to the base year stock. Worker death and retirement rates differ by age, and so age-specific rates must be built up for worker age groups, 15-19, 20-24, and so on. The rates are applied to the work force divided into five-year age groups, just as in demographic projections. Since each occupation has a characteristic age structure the wastage can be computed for occupations within sectors and then related to education. To save time, analysts sometimes apply one rate across all age and occupation groups in the work force. (A common rate is three percent cumulative annually).

Outflows or wastage from supply forecasts are estimated and applied the same way, although rates will be lower because the newly entering workers are younger. Outflow rates are made up of two main components: mortality rates and retirement rates. Various tables of work force wastage rates have been built up by the UN and the ILO, although each country will differ and planners often build their own just as demographers build their own model life tables for countries (see Chapter 3). In carrying out projections for areas within a country (labor market areas), the estimation of rates may be quite difficult because of internal migration from one labor market area to another. Transfers from one sector to another or one occupation to another also occur.

5.4 PROJECTING SUPPLY AND COMPARISON OF SUPPLY WITH DEMAND

To compare the requirements, or demand, with the supply or output of educational and training institutions requires a projection of the output of the education-training system. Supply is projected by applying enrollment flow models to the education and training system, as discussed in previous sections. The projected output, or supply, of the system with wastage taken into account is compared to the projected requirements to yield an estimate of deficits or surpluses in the supply of educated or trained people, as shown in Figure 5.7 on the next page.

Figure 5.7

Comparison of Economic Demand and Educational Supply in Chilean Manpower Forecast

	Number Completing, 1960-1970		
	Economic demand[a]	Total supply[b]	Difference (demand-supply)
18 Grades (Univ)	12,213	6,440	+ 5,773 (deficit)
17 Grades (Univ)	38,643	25,920	+ 12,723
12 Grades (All Sec)	117,952	181,570	- 63,618
6 Grades	497,539	1,053,440	- 555,901

[a] From Table 10.8.
[b] 1962 Graduate multiplied by 10 (note problems in this method)
[c] All Secondary except Feminine Technical

Note: In Figure 5.7 the supply was forecast by such a crude method that it did not seem worth while to take out wastage.

5.5 EMPLOYMENT FORECASTING BY THE ALTERNATE METHOD

Alternate methods of manpower requirements forecasting use curve fitting. A curve is fitted to past series data and a trend extrapolated to the future. In the Dominican case, the fit is to an exponential form:

Figure 5.8

Dominican Examples: Manufacturing Sector

Industry	Occupation		1985 extrapolation
Sugar	Engineers	Number 1970 = 22	$22\left[e^{.55(.049)} + .043\right]^{15}$
			$= 22(1.033)^{15} = 22(2.759) = 61$
Textiles	Managers	Number 1970 = 62	$62\left[e^{.30(.10)} + .007\right]^{15}$
			$= 62(1.037)^{15} = 62(1.7246) = 107$

In the Dominican case, different methods for projecting occupation and employment were used for different sectors:

1. For construction, transport, electricity/gas/water, the product for the sectors was first projected to the target year 1985, on

the basis of a curve fit to past growth trends and a modification to reflect present and future expansion projects and plans.

2. For the government services of education and health, employment and occupational distributions were based on population forecasts and service norms and on staffing ratios.

3. For mining, where the sector is dominated by a few large modern enterprises, the expansion plans, including detailed manpower requirement forecasts, were incorporated directly into the general sectoral manpower requirements forecasts. Manning table data were also directly used.

5.6 OTHER METHODOLOGIES

There are as many other different methods or modifications of the basic methodology as there are planners and analysts at work on macro-level manpower forecasting.

1. Curve Fitting and Extrapolation

At the base of most alternative methods, e.g., formula 1) used in the Dominican case, are models and methods for curve fitting and extrapolation of one or more of the major components used in manpower forecasting: (1) Product, (2) Productivity, or (3) Employment.

$$P_n = P_x (1 + r)^t$$

Here product in year n is extrapolated on the assumption that it will increase at rate r compounded annually. An exponential curve, as in the Dominican formula, is simply the continuous function version of the compound interest formula. The same curve functions can be assumed, or fitted, in order to show increase in productivity, employment, employment in any given sector, or employment in any given occupation. Most spreadsheet packages provide functions for fitting curves.

Estimates are also made on the basis of regression analysis, simple (single independent or "predictor" variable) or multiple (more than one independent) as shown on the next page:

Simple Regression: $P = Pa + bX$

A common practice is to fit the output or dependent variable (say, product) to a series of annual outputs in past years (where X stands for the year). With the constant a estimated, and with the parameter b coefficient, the analyst can substitute a future year X and extrapolate the product estimate.

Multiple Regression: $P = a + b1X1 + b2X2 + \ldots bnXn$

Here the X1, X2, etc. independent variables are other variables that are related to the dependent variable product, for example, investment in capital, capital/output ratios, etc. They can be used to estimate the parameters for future year forecasts.

2. Beyond Curve Fitting and Extrapolation

Many economic forecasts use regression analysis based on single equation models, and on one-stage least squares. Curve fits or extrapolations from serial data have a similar underlying mathematical logic and similar limitations in the matter of dealing with simultaneity among variables (where the solution of one variable affects the values of another variable). The packages Forecast Master, or Smart Forecast, offer the analyst a wide variety of curve fits and tests that can be applied to serial data used in economic projections for planning.

Planning has also been based on sets of linear difference equations where the solution or assignment of value to one variable leads successively to the solution of other variables over time. Tinbergen pioneered in the development of this form of econometric modeling (1958); Tinbergen and Bos (1965) developed linear difference equation models for estimating and projecting manpower requirements and educational outputs in terms of economic growth. The path was simple:

i. Manpower with a given level of education was expressed as a linear function (by a coefficient of proportionality) of growth in product (as set in economic plan targets).

ii. The numbers in manpower/education-level categories were traced from step i) back to required levels of flow and output in the education, given: the numbers in the work force with a given level of education in a previous period; the numbers

required to enter the work force to replace educated workers dying or retiring; and the numbers of educated workers needed for teachers (based on teacher/pupil ratios).

iii. Once base-period values of educated people in the work force were set, plan targets in economic growth set, staffing and technological ratios set (teacher/pupil ratios), and wastage rates estimated from past data, the analyst could substitute values and solve his way through the equation system in terms of prior period values.

One problem is that interaction and simultaneity in the system were not adequately reflected in any single equation or single-stage analytic models, and planners moved to get around this with econometric models and dynamic programming models.

3. Econometric Models: A Glimpse Beyond

Econometric models are sets of linked linear equations that are solved simultaneously in order to obtain the value of an endogenous variable (solved within the model system) from the value of an exogenous variable(s) that has been previously estimated or assigned a value on the basis of other knowledge.

For example, enrollment can be estimated on the basis of economic demand (taken from plan growth targets), and social demand estimated on the basis of a regression equation where the independent variables are X1 (population in school age), X2 (literacy level), X3 (income level) . . . The model is stated in terms of <u>structural</u> equations which represent the real world system modeled. The equations are solved so that the result is a <u>reduced form</u> equation in which the endogenous variable (variable of interest to be solved for) is stated as a function of exogenous variables that can be estimated from real-world data by regression.

The regression analysis is performed in the first stage of the model solution, and the values substituted into the reduced form equations and a second-stage least-squares run to estimate the value of the endogenous variable. The endogenous variable value can then be used in a projection of future values in terms of changes in independent variables or estimators.

There are some statistical problems in using the estimates, the main ones being multicollinearity (variables in the predictor set are highly correlated with each other), heteroscedasticity (error terms correlated), and autocorrelation (correlation in one variable value over time series). The magnitude and effect of these errors in time series data on which econometric models are based can be estimated using Forecast Master which has Durbin-Watson, A/C, and other tests.

4. System Dynamics Models

The term "System Dynamics" is associated with Jay W. Forrester (1965, 1968, 1971) who applied the models and concepts to industrial engineering (where the level of the variables over time was the point of interest). Forrester attempted to get around the problem of simultaneity in equation sets by modeling the system as a set of integral equations in continuous time, where the interval dt was taken so small as to approach simultaneity and the values are solved for serially but with simultaneity simulated, using numerical approximations (Euler's method) from calculus to evaluate the series value. System dynamics models were programmed for large computers, using Dynamo (Pugh, 1981), and more recently for PCs using STELLA (Richmond, 1985, 1988).

Development projects require models for linking macro-level models and methods of planning and policy analysis to the micro-level research and evaluation of teaching/learning required for program-level design and planning.

Comprehensive models may link all demographic, economic, and school component forecasts within a single systematically related structure or system in which process in one part of the system affects process in another. At any given time, however, the planner may be interested in only one component of the forecast model. In theory, synoptic planning favors comprehensive models in which all major components are assembled and linked; but practice may favor forecasts of one component at a time, either because of limited interest or need, or limited data and time.

Exercise

USING SMART FORECASTS II

5.7 EXERCISE OBJECTIVES

The objectives of this exercise are:

1. To present the concept of time series analysis and its use for educational plannning.

2. To demonstrate the purpose and capabilities of the Smart Forecasts II program.

5.8 INTRODUCTION

Smart Forecasts II (SFII) is a PC software package that simplifies time series analysis and forecasting. SFII can be used to apply time series analysis to data series:

To analyze relationships among variables, e.g. using regression.

To forecast the values of variables.

SFII offers clear menus that call up pre-programmed procedures:

To describe and explore time series data (variable measurements recorded at regular intervals, days, years, in the past).

To use the analysis of trends or regularities in the past (a) to forecast future values and (b) to test the forecasts.

SFII has two major advantages for the user. First, the user need not deal with analysis and forecasting models, methods and algorithms (the solution procedures), but only with menu choices that call them up. In fact the user need not select from the five basic forecasting algorithms, which will be discussed briefly, but only select AUTOMATIC on the Forecast sub menu and the program will run the forecast,

Forecast sub menu and the program will run the forecast,

test the fit, and select the best solution. Thus, much of the analysis is automatically done by the computer. The user need not be an expert in time series analysis and forecasting.

Secondly, statisticians who wish to choose a specific method on their own may bypass AUTOMATIC and choose any one of five "specific" methods, such as Multivariate Regression. SFII will run univariate forecasts, forecasts of the future values of a single variable, and multivariate forecasts, forecasts based on series of values of multiple variables. Usually multivariate forecasting uses some form of regression in order to forecast the values of one dependent variable, using one or more independent variables.

SFII is used to analyze patterns in the form of:

Trends: Regular changes (increase/decrease) in values over time

Cycles: Recurrent patterns of highs and lows

Seasonal patterns: Changes that can be related to time periods

Trends, cycles and seasonalities are what time series analysts deal with in forecasting, with SFII, or any forecast package.

Analysts who know the trends and data patterns in their own field or business can build this knowledge into the forecasts by a procedure called EYEBALL which enables the user to build hunches, experience, and intuitions into shaping the curve. Knowledge about fluctuations and discontinuities that affect trends, or seasonal variation, like purchase of beach wear in the summer, can be used for seasonal adjustments or smoothing.

Resulting forecasts can be studied with statistical tests and table comparisons, and schematized with graphics. The adjustments and retrials are painless to carry out. What-if trials can be run more systematically and comparisons shown more graphically than with spreadsheets.

Importing and exporting series data from and to Lotus are easy procedures. By converting to ASCII and DIF file form, data can be built into other packages (or in DOS EDLIN) to be interchanged or down loaded from larger computer systems. Although the process is simple, a few interesting features should be noted.

1. A General Note on Time Series Forecasting

Planners apply time series analysis and regression to data measured over time in the past in order to estimate parameter values that describe patterns or trends that can be used to forecast future values of the data. Data can be shown as points on a flat surface, a plane, where one axis (Y) measures the value of the variable and the other axis (Xt) is time in months, or years.

A typical linear time series is: $Y = a + b\,Xt$
Here Y is the value, Xt is the time the value is measured, e.g. in years, and **a** and **b** are parameters that shape the line. The **a** parameter shows where the line cuts the Y axis, and the **b** shows the slope and direction (up/down) of the line. The parameter values a and b are estimated by fitting a line to a set of data points based on past values of the variable (Y's).

2. A Technical Note on Curve Fitting for Forecasting

<u>Fitting</u> means selecting a line that runs through the patch of data points so as to minimize the distance of all the data points from the line. Mathematically this is done by a line in which the squared distances, or deviations of the points from the line are <u>minimum</u>. When these deviations are <u>minimized</u> in the calculus, i.e. differentiated partially with respect to **a** and **b** and set equal to zero, it results in two equations that can be solved simultaneously to get **a** and **b** in the line, $Y = a + bX$ that <u>best fits</u> the data points. Once the a and b values are known, the analyst can solve for Y values in terms of any X value. Since X is time in years, analysts can <u>extrapolate</u>, or get the value of the Y variable for any year X, for any time in the future. This is what forecasting does using curve fitting. In this case the curve form is linear, but other forms, e.g. exponential and logistic, can be used.

To summarize:

1. The analyst has a set or series of past data on a variable. In time series, the measures are made at a regular interval (like a year or quarter) over time.

2. The analyst plots these points on a plane, using the Y axis for the values of the variable, and the X axis for the time period (years, months, quarters ...)

3. The analyst fits a linear curve form to the data points to obtain the best fit line. The best fit line is a line determined by the parameters **a** and **b** that minimizes the sum of the squared distances of the points from the line.

4. In fitting this line, to minimize the distances, the values of a and b are derived to fit a given set of data points.

5. Using a and b values, the analyst can extrapolate, or get any Y value for any X value. To go back to the plane, this means extending the line. In practical terms, it means getting Y values for future years, since these are the X values.

6. This is one form--the most basic--of forecasting.

7. The same linear model can be used to get one variable value, X1 from other variable values X2, X3... in regression based models.

8. Thus the one simple, linear model $Y = a + bX$ or variants with multiple X's can be used for the main tasks of planning, researching variable values and forecasting variable values.

The first use, which focuses on evaluating and researching values of educational variables, will be discussed later. Here the task is forecasting. SFII is a tool designed to help planners make forecasts.

Planners deal with trend analysis and extrapolation in order to forecast the basic variables of social, economic and educational planning. These variables include population, enrollment, and employment. The same basic form and concepts apply to the basic liner model used for:

1. Time series analysis/curve fitting/forecasting, as in the equation $Y = a + b X_t$ where the value of the variable is on the Y axis, and time on X.

2. Simple linear regression, where the value of the dependent or endogenous variable is on the Y axis and value of the independent or exogenous variable is on the X axis.

The two basic use forms of the simple linear equation model are:

1. Time Series: $Y_t = a + Y_{(t-1)} + e$

Mathematically these are same form. They fit a line to a set of data points so that the distance of the data points from the line is minimized, i.e. the sum of the squared distances from line to data points is minimized. Hence the fit procedure is called least squares. The fitting procedure gives two simple linear equations called "normal equations". (Normal equations derived from least squares and the set of data points can be solved simultaneously to get the values of the parameters a and b). Once the values of a and b are known for any set of data points, they can be put back into the equation:

$$Y(t+1) = a + b\,Y(t)$$

This gives the future value of Y (there is also some error, e). Note that there can be more than one independent variable, as in multiple regression:

$$Y_t = a + b_1X_1 + b_2X_2 + b_2X_3 + \ldots$$

Time series and regression are based on the same linear model.

5.9 FIRST STEPS IN USING SFII

Forecasting is the purpose of SFII, but this is not the way to begin. As Figure 5.9 on the next page shows, there must be variable values to work with. Hence, the first step must be to load the work table with data.

The first time user will generally:

1. Switch on the computer, change to the appropriate directory (SF in the BRIDGES lab computers).

2. Type SF to activate Smart Forecasts II, then press ENTER.

3. Select HELP on main menu to get overview of all SFII Commands.

The Main menu, also called the Edit Mode menu, is shown in Figure 5.9.

The Help Overview menu is shown in Figure 5.10.

4. For details on any group of commands, or for specific commands, type HELP followed by command, e.g. HELP Forecast, to get an explanation of the Forecast command. Help may be accessed from <u>any</u> mode in SFII.

HELP Forecast, to get an explanation of the Forecast command. Help may be accessed from <u>any</u> mode in SFII.

Figure 5.9

The First Screen of SFII

```
SmartForecasts II     (C) 1985, 1986, 1987  Smart Software, Inc.

                    ............. ............. .............
                       V1            V2            V3
                    ---------------------------------------------
............. C1 :        :             :             :
............. C2 :        :             :             :
............. C3 :        :             :             :
............. C4 :        :             :             :
............. C5 :        :             :             :
............. C6 :        :             :             :
............. C7 :        :             :             :
............. C8 :        :             :             :
............. C9 :        :             :             :
............. C10 :       :             :             :
............. C11 :       :             :             :
............. C12 :       :             :             :
                    ---------------------------------------------

EDIT MODE : Read

For help, type HELP EDIT or HELP followed by one of these commands:

CLEar  DEFine  DELete  EXPlore  FORecast  HELp  INSert   LABel  LISt
DOPs   QUIt    READ    REPlace  SAVe      SEE   SEQuence TITle  VARiables
```

1. The Main, or Edit, Menu is below an empty worktable because no values have been READ in yet.

2. The user types <u>READ</u> to read in values of a variable.

```
SmartForecasts II     (C) 1985, 1986, 1987  Smart Software, Inc.

                    ............. ............. .............
                       V1            V2            V3
                    ---------------------------------------------
............. C1 :        :             :             :
............. C2 :        :             :             :
............. C3 :        :             :             :
............. C4 :        :             :             :
............. C5 :        :             :             :
............. C6 :        :             :
............. C7 :        :             :
............. C8 :        :             :
............. C9 :        :             :
............. C10 :       :             :             :
............. C11 :       :             :             :
............. C12 :       :             :             :
                    ---------------------------------------------

What do you want to READ? (Press W or L or A or D or T or Esc ):
Press  W  to read the  SFII WORKSHEET: ENROLLMENTS
Press  L  to read a    Lotus wrk  data file.
Press  A  to read an   ASCII      data file.
Press  D  to read a    DIF        data file.
Press  T  to read the  TUTORIAL   data table.  To cancel, press Esc .
```

3. User may select five different types of data files to read.

Figure 5.10

Help Overview in SFII

```
                    Overview of SmartForecasts II

EDIT mode: to edit data tables.              Commands are:

read      clear     label     insert         variables  help  quit
save      replace   sequence  delete         list       oops  explore
          define    title                    see              forecast

EXPLORE mode: to explore your data.          Commands are:

describe  scatter   decompose correlate      variables  help  quit
          timeplot                                           edit
                                                             forecast

FORECAST mode: to make forecasts.            Commands are:

automatic simple    single    multiseries    variables  help  quit
          linear    double    eyeball                       edit
                    winters   regress                       explore


For more help with OVERVIEW, press  Enter ; otherwise, press  Esc ;
```

5. After reviewing the HELP facilities in SFII, you are ready to load data into the SFII worksheet.

6. With data in the worksheet you may begin to EXPLORE the data and to execute FORECASTs.

5.10 PREPARATION FOR FORECASTING WITH SFII

1. Loading the SFII Worksheet with Data

Although the SFII worktable can be loaded directly with data and can be used to write labels and titles, users familiar with Lotus 1-2-3 will find it easier to structure their data tables in Lotus and use the SFII READ command to import data into the SFII worktable. However, <u>all data must be READ (imported) into the SFII work table columns in one move.</u> The work table can not be built up successively, variable by variable using the READ command. Each time a

move. The work table can not be built up successively, variable by variable using the READ command. Each time a

new file of data is READ, the previous worktable is first cleared of all data. Thus to load the SFII worktable with a multivariate (multiple column) series, all columns should be set up in Lotus first. Then READ them into SFII as one set.

To READ a Lotus file into the SFII work table, SFII requires that the Lotus spreadsheet be arranged in one of six specified formats. Most users will find easiest to use one of the following formats:

Example 1

Fully labeled; title in cell A1; variables in rows.
READ columnwise into SFII.

	A	B	C	D
1	Enrollments	GRADE1	GRADE2	GRADE3
2	1976-77	766088	820319	691171
3	1977-78	786005	837675	710200
4	1978-79	793572	852252	718246

Example 2

Fully labeled; title in cell A1; variables in columns.
READ rowwise into SFII.

	A	B	C	D
1	Enrollments	1976-77	1977-78	1978-79
2	GRADE1	766088	786005	793572
3	GRADE2	820319	837675	710200
4	GRADE3	691171	710200	718246

It is also possible to READ only a portion of a large Lotus spreadsheet into the SFII worktable by extracting the appropriate data range from the Lotus spreadsheet using the / FILE EXTRACT in Lotus. But the portion you extract must be arranged as in the examples above. SFII does not READ the first column or the first row as data.

It is also possible to omit titles and labels from the Lotus spreadsheet and to TITLE the work table and LABEL the variables (top of column) by using SFII commands (shown on Main, Edit Menu). To do so, leave the first column and

With the READ command, data may also be imported in other forms, such as ASCII, DIF, etc. (See Figure 5.9.)

[NOTE: In Table 5.9, parts 2 and 3 show the beginnings of the sequence of READ commands. A detailed walk through of the SFII READ process can be found in Annex A.]

A variety of editing commands are listed on the main menu for adjusting or changing variables:

a. REPLACE is the data edit command for changing any values.

b. INSERT puts in variable columns.

c. DELETE takes them out.

In the SFII worktable, variables are displayed in the columns and cases along the rows. Twelve rows and three columns are displayed on the screen at any one time. To see any value in the series outside the displayed range, use command SEE, followed by sequence number, e.g "SEE C88". The work table then displays 12 sequential values centering on the value chosen, e.g. from C83 to C94. Figure 5.11 on the next page gives an example of the SEE command. The user can also use SEE to move across columns. "SEE V6" will go to column 6.

After the data is read in, it is time to use EXPLORE.

2. Building New Variables and Forecasting Structures

DEFINE can be used with the SFII worktable to build new variables from composites of old variables, to add, subtract and multiply variables, and to write variables in new formats or mathematical forms (absolutes, integers, rounded). DEFINE can also be used to build in time series structures and seasonal leads and lags, when one variable is shown as leading or lagging behind another variable, through the use of CROSS CORRELATION in EXPLORE.

Figure 5.11

Use of the SEE Command

```
SmartForecasts II     (C) 1985, 1986, 1987  Smart Software, Inc.

                    ............  ............  ............
                       V1            V2            V3
                    ---------------------------------------------
............ C1  :    112.030       :             :
............ C2  :    113.670       :             :
............ C3  :    114.930       :             :
............ C4  :    115.930       :             :
............ C5  :    117.130       :             :
............ C6  :    118.200       :             :
............ C7  :    119.700       :             :
............ C8  :    121.900       :             :
............ C9  :    123.500       :             :
............ C10 :    124.530       :             :
............ C11 :    125.800       :             :
............ C12 :    127.070       :             :
                    ---------------------------------------------

EDIT MODE : see c88

For help, type  HELP EDIT  or  HELP  followed by one of these commands:

CLEar  DEFine  DELete  EXPlore  FORecast  HELp  INSert    LABel  LISt
OOPs   QUIt    READ    REPlace  SAVe      SEE   SEQuence  TITle  VARiables

     SmartForecasts II     (C) 1985, 1986, 1987  Smart Software, Inc.

                    ............  ............  ............
                       V1            V2            V3
                    ---------------------------------------------
............ C83 :    212.170       :             :
............ C84 :    215.500       :             :
............ C85 :    219.370       :             :
............ C86 :    224.530       :             :
............ C87 :    228.100       :             :
............ C88 :    229.930       :             :
............ C89 :    234.730       :             :
............ C90 :    238.730       :             :
............ C91 :    243.530       :             :
............ C92 :    249.470       :             :
............ C93 :    254.600       :             :
............ C94 :    257.700       :             :
                    ---------------------------------------------

EDIT MODE : see c110

For help, type  HELP EDIT  or  HELP  followed by one of these commands:

CLEar  DEFine  DELete  EXPlore  FORecast  HELp  INSert    LABel  LISt
OOPs   QUIt    READ    REPlace  SAVe      SEE   SEQuence  TITle  VARiables
```

3. Exploring the Data Before Forecasting

A prudent analyst explores the data before forecasting, first visually in the Edit Mode with the SEE command. The command VARIABLES will show a chart of the variables in the work table. DESCRIBE will show common descriptive statistics (mean, median, mode and standard deviation) and a histogram picture of distribution. A more systematic exploration can be made using the EXPLORE command. EXPLORE and DESCRIBE are explained in Figure 5.12 on the next page. Results of DESCRIBE for V1 are shown Figure 5.13 (page 41).

One unique feature of EXPLORE is the choice of TIMEPLOT which will graph a univariate series over time and reveal the shape of the curve. For example, the variable graphed with TIMEPLOT in Graph 5.1 (page 42) shows exponential increase over time which can be used to select a form from the FORECAST menu. SCATTER shown in Graph 5.2 (page 43) helps explore relationships between two variables.

CROSSCORRELATION in Graph 5.3 (bottom of page 43) shows lead and lag relationships between a pair of variables. Knowing this, the analyst can forecast the value of one variable from the past value of another variable, and build this with DEFINE.

Figure 5.12

Explanation of EXPLORE and DESCRIBE

Help for: EXPLORE

Use EXPLORE mode to analyze and graph your data. Commands are:

CORRELATE..display autocorrelation or crosscorrelation plots.
DECOMPOSE..break a series into components and seasonally-adjust it.
DESCRIBE...get descriptive statistics and histograms for up to 3 variables.
EDIT.....switch from EXPLORE mode to EDIT mode.
FORECAST...switch from EXPLORE mode to FORECAST mode.
HELP.....temporarily switch to HELP mode.
QUIT.....end a session with SmartForecasts II.
SCATTER....display a scattergram plotting one variable versus another.
TIMEPLOT...display timeplots or linegrams for up to 3 variables.
VARIABLES..display the variables and variable labels in the data table.

HELP MODE : Enter a command name for which you want help:
(Press Enter to get an overview. Press Esc to end HELP.)

Help for: DESCRIBE

EXPLANATION:

DESCRIBE computes descriptive statistics and displays histograms for up to 3 variables. The descriptive statistics computed are: mean, median, standard deviation, maximum, minimum, and count of cases with non-missing data. A histogram is a bar graph that shows how the data values are distributed between the minimum and maximum.

EXAMPLES:

To describe V1:	DES V1<Enter>
To describe V1 and V3:	DES V1 V3<Enter>
To describe V1, V3 and V2:	DES V1 V3 V2<Enter>

HELP MODE : Enter a command name for which you want help:
(Press Enter to get an overview. Press Esc to end HELP.)

Figure 5.13

EXPLORE Using DESCRIBE

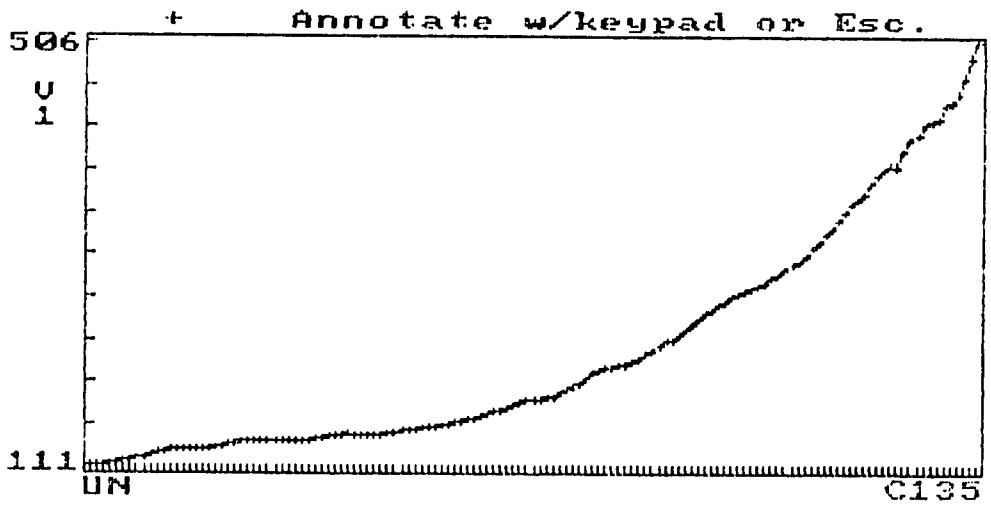
	IV1:		
Mean.....	219.324		
Median.....	173.230		
Std.Deviation..	103.412		
Maximum.....	505.200		
Minimum.....	112.030		
Number of cases	134		
Histogram.....			

```
EXPLORE MODE :

For help, type HELP EXPLORE or HELP followed by one of these commands:

CORrelate  DECompose  DEScribe  EDIt      FORecast
HELp       QUIt       SCAtter   TIMeplot  VARiables
```

Graph 5.1

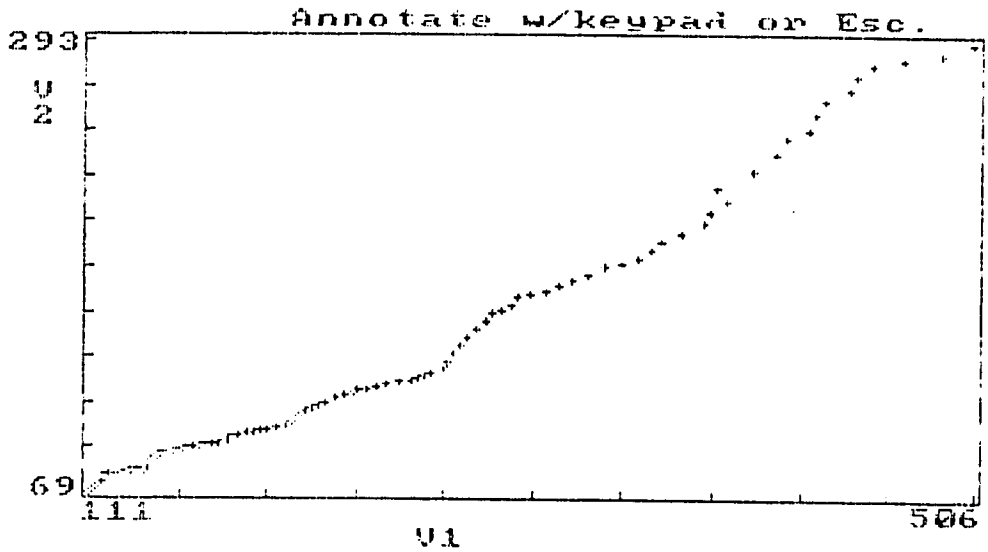
TIMEPLOT Used to EXPLORE Trends
in One Variable

```
-----------------------------------------------------------------------------
EXPLORE MODE : TIMEPLOT V1        Xt          Xt-1

For help, type HELP EXPLORE or HELP followed by one of these commands:

CORrelate  DECompose  DEScribe  EDIt      FORecast
HELp       QUIt       SCAtter   TIMeplot  VARiables
```



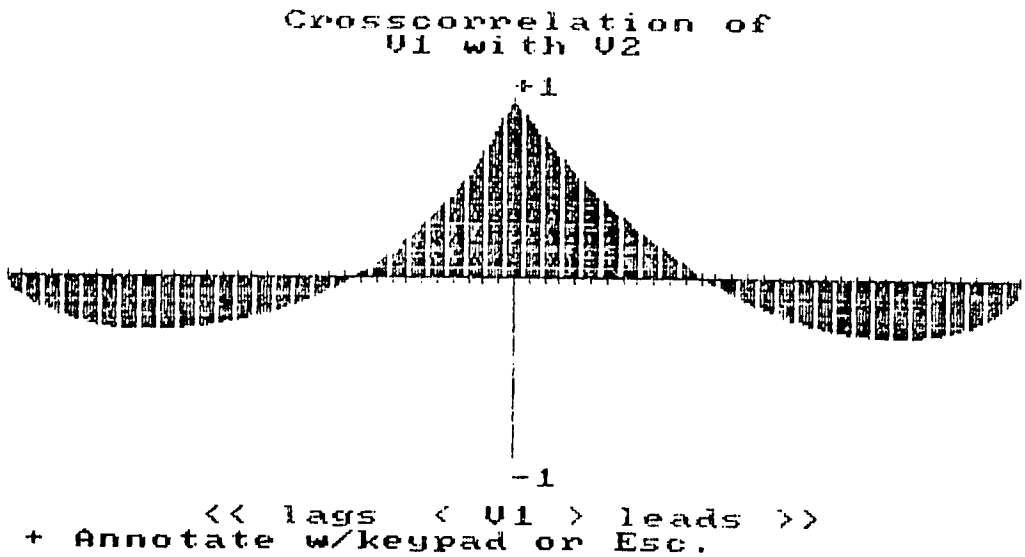
Graph 5.2

SCATTER Used to Explore
Bivariate Relations (Two Variables



Graph 5.3

CROSSCORRELATION to Explore Two Variables to Identify
Variables that Lead or Lag Another Variable



5.11 FORECASTING WITH SFII

In SFII, all forecasting methods follow the models in the two equations mentioned earlier:

1. $Y_t = a + Y_{(t-1)} + e$ (Time Series)

2. $Y = a + bX + e$ (Linear Regression)

For equation 1 the procedures are used mainly in fitting univariant time series data. The five algorithms used by the AUTOMATIC command in the Forecast Mode use variants of equation 1. AUTOMATIC chooses from the following algorithms:

1. Single exponential smoothing
2. Double exponential smoothing
3. Simple moving average
4. Linear moving average
5. Winter's multiplicative exponential smoothing

The user can also go beyond AUTOMATIC and choose REGRESS, which uses multivariate regression with up to 20 predictors. For multivariate analysis, the basic five algorithms applied to univariate series and REGRESS use both the linear model and patterns of past variation in the variables to predict future values. In univariate (one variable) forecasts, the past data values of the variable are analyzed to get a trend. Then, when the parameters which describe this trend are estimated, they can be used to extrapolate, or forecast, future values.

The major difference in the five algorithms listed above, which all apply to single variable forecasts, is that seasonal patterns are smoothed or averaged out to improve the forecast based on the overall trend. Interest is in four main patterns of variation:

Seasonal
Cyclical
Trend
Irregular

In SFII the main interest is in Trend. SFII uses "smoothing" or "averaging" on the five variants above to account for seasonal patterns and irregularities. It uses, as indicated, two curve forms, exponential and linear for

1. AUTOMATIC Forecasts

A normal first choice in FORECAST is AUTOMATIC which runs the five basic algorithms on the data and selects the one that fits best. AUTOMATIC is chosen and run for Variable One in Figure 5.14 below. Results are shown in Graph 5.4 on the next page. The selection is a Linear Moving Average of sets of five values in the univariate time series. Note it is exponential in form.

Figure 5.15 (below Graph 5.4 on the next page) shows that there are many different forms of output that can be used in AUTOMATIC. The most common and useful forms are Table and Graph. [Note: SFII also shows how the five methods rank in the tournament, using this set of data.]

Figure 5.14

A Typical SFII Worktable of Variables (Two are shown, Three are not shown)

A Set of Univariate Economic Time Series Variables

		V1	V2	V3
UN.....	C1	111.980	69.850	267.005
.....	C2	112.030	70.650	267.600
.....	C3	113.670	71.190	277.100
.....	C4	114.930	72.480	294.800
.....	C5	115.930	73.960	306.300
.....	C6	117.130	76.980	320.400
.....	C7	118.200	77.980	328.300
.....	C8	119.700	77.980	335.000
.....	C9	121.900	77.710	339.200
.....	C10	123.500	77.710	341.900
.....	C11	124.530	77.670	342.100
.....	C12	125.800	78.830	347.800

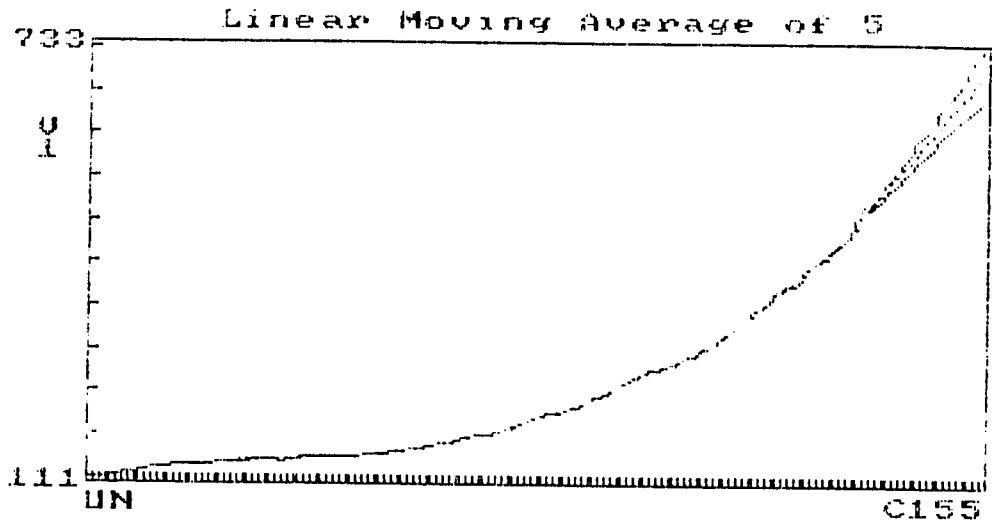
EDIT MODE :

For help, type HELP EDIT or HELP followed by one of these commands:

CLEar DEFine DELete EXPlore FORecast HELp INSert LABel LISt
OOPs QUIt READ REPlace SAVe SEE SEQuence TITle VARiables

Graph 5.4

Forecast is chosen and AUTOMATIC Within Forecast



NOTE : In this case Automatic selects Linear Moving Average of Sets of Five Variable Values as Best Fit Curve

Figure 5.15

Various Forms of Output from an AUTOMATIC Forecast

Special Function key options for the AUTOMATIC command:

F1 = see graph of forecasts	F2 = save forecasts + smoothed data
F3 = see table of forecasts	F4 = see tournament rankings
F5 = print table of forecasts	F6 = save upper + lower forecast limits
F7 = save forecasts only	F8 = annotate (if graph is displayed)
F9 = save forecasts + history	F10:
Alt F1 = change % level of forecast interval	End = repeat forecast command for next variable

Esc = return to FORECAST mode

2. Building Experience and Knowledge into Forecasts

The ease of doing automated forecasts is not the main feature of SFII. Quite the contrary, the most impressive feature of SFII is that it encourages users with special knowledge of the field in which they are forecasting with creative and artistic notions about how the forecast should look. The users can build these hypotheses into the forecast using a function called EYEBALL. This is a submode of FORECAST. It enables the user to make manual adjustments to the shape of the forecast curve, to take into account special knowledge of events or discontinuities that might shape trends. For example, some annual or seasonal event that drives trends up or down may reoccur.

To begin work with EYEBALL:

1. Call SFII, get main or EDIT menu screen.

2. Select READ, to read in one or more time series variables as data.

3. EXPLORE can then be used to examine the form of the data series, with DESCRIBE, TIMEPLOT, SCATTER (if working on more that one variable).

4. With some idea about trends or seasonal patterns that will shape the forecast, the user generally opts for FORECAST to try forecasting the variables(s).

5. In FORECAST mode, a reasonable first step is AUTOMATIC to let the program select the best of the five forecast methods described above, based on the data that was read in.

6. EYEBALL is a sub mode within FORECAST. EYEBALL is used to adjust the best forecast selected by the program in AUTOMATIC. For example, typing the command "Eye V1" permits the user to examine and change variable 1 forecast.

7. Type in number of forecast periods, e.g. "10" and press ENTER. This brings a graphic on screen with two forecasts, labeled "A" and "B". (The A forecast is a standard against which the B forecast will be adjusted by EYEBALL).

standard against which the B forecast will be adjusted by EYEBALL).

At the bottom of the screen, the function keys guide the following steps:

a. <F4>, function key four, calls EYEBALL.

b. <F2> calls up the best forecast under AUTOMATIC as a new B that will be adjusted with EYEBALL to improve on A.

c. Type in "V1" to give automatic Forecast B a variable designation.

d. <F1> allows manual adjustments to B on screen.

e. Use "up" arrow keys to move up to curve line on graph. Use "right" or "left" arrow keys to move along line to next or prior values on the line. [NOTE: This movement increases or decreases the forecast value on the screen at the cursor.]

f. <Ins> puts in the changed forecast value at the cursor, i.e. user shapes curve according to his hunches or notions.

g. Press ESC key when hunches or ideas about the shape of the forecast are inserted on graph, to end and save new values.

h. Press <F4> to switch A and B labels. Now the adjusted curve B is the standard or preferred forecast A.

i. ESC ends the EYEBAL submode of FORECAST adjustment session.

j. <F7> saves the new A as the forecast.

k. Type "V4" to enter the values of A as variable on worksheet. SAVE as a worksheet in SFII or SAVE as a separate file on any drive, or SAVE to a Lotus or DIF file.

l. ESC ends the EYEBALL session.

3. OOPS Command

Just as EYEBALL makes SFII more adjustable and flexible, the command OOPS on the main menu makes it more reversible. OOPS will undo any command that has been made immediately before, with no other intervening command. When you experiment with curve fitting, because curve fitting is

4. Multivariate Forecasts: Regression

Figure 5.16 below shows data arranged for a multivariate forecast, in this case REGRESSION. Graph 5.5 on the next page shows the results. Below are statistical tests of the results. In the graph, the fit looks close, and the R Square shows a very high percent of variance explained in V1 by variables 2, 3, 4, and 5, almost 100 percent. If too many variables are used, the forecast may not be efficient. Note that if enough predictors are used, almost everything about a dependent variable can be explained and forecast.

The Durbin Watson statistic shows (the optimal average value should be about 2.00) that the residual errors are correlated. In some cases, this would suggest a biased series. Durbin Watson runs from 0 to 4, and the ideal centers around 2. If the value falls at the extremes 0 or 4, then REGRESSION in which it occurs may show a biased historical relationship, although this may still make a reasonable set for forecasting. Correlation among the errors can be examined in the plot of RESIDUALS and <F2> will display a plot of the residual errors. This should show that residuals are randomly scattered rather than in a pattern, which shows bias. Pressing <F6> will table the best and worst fit values in a REGRESSION run.

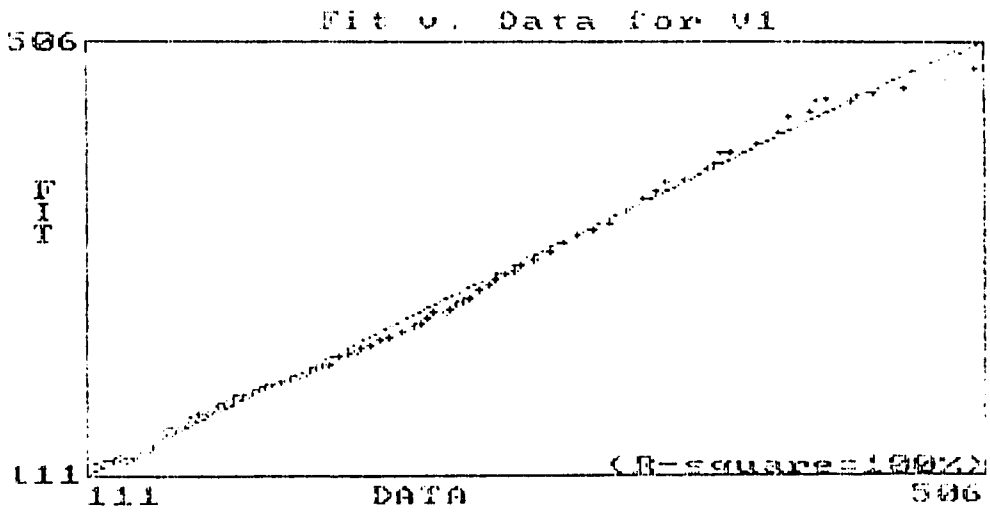
Figure 5.16

Multivariate Forecast Setup
(The Choice will be to Predict V1 from V2 V3 V4 V5)

```
A Set of Univariate Economic Time Series Variables
List of variables:
------------------
V1 :(in use)     V16:          V31:          V46:
V2 :(in use)     V17:          V32:          V47:
V3 :(in use)     V18:          V33:          V48:
V4 :(in use)     V19:          V34:          V49:
V5 :(in use)     V20:          V35:          V50:
V6 :             V21:          V36:          V51:
V7 :             V22:          V37:          V52:
V8 :             V23:          V38:          V53:
V9 :             V24:          V39:          V54:
V10:             V25:          V40:          V55:
V11:             V26:          V41:          V56:
V12:             V27:          V42:          V57:
V13:             V28:          V43:          V58:
V14:             V29:          V44:          V59:
V15:             V30:          V45:          V60:
-------------------------------------------------------------------------------
FORECAST MODE :

For help, type HELP FORECAST or HELP followed by one of these commands:

AUTomatic  DOUble EDIT   EXPlore EYEball HELp     LINear
MULtiseries QUIt   REGress SIMple  SINgle  VARiables WINters
```

Graph 5.5

Results of Multivariate Forecast: REGRESSION



Regression fit to V1
Analysis uses 135 cases from C1 ON to C135.

PREDICTOR	COEFFICIENT	STD.ERROR	t-VALUE	P(2-tail)
(constant)	97.494			
V2	-0.393	0.089	-4.42	.01
V3	0.148	0.006	26.24	<.01
V4	-0.501	0.439	-1.14	n.s.
V5	1.822	0.548	3.32	.01

R-square = 99.6% Adjusted R-square = 99.6%
# of cases = 135 F = 8732.2 d.f.=(4,130)
Standard Error of Estimate = 6.395 Durbin-Watson = 0.19

F1=fit F2=resid F3=coef F4=fcst F5=prnt
F7=save.fit F8=save.resid F6=cases Esc

V1", SFII presents an autocorrelation plot, or a correlogram, which shows the correlation of the variable with itself over time. This shows correlation of one value of the variable with its own values, earlier or later. CORRELATEcan thus be used to identify seasonal patterns, for example if sales always lagged in the month of January over a period of years.

CROSSCORRELATION, previously mentioned, gives a graph of one variable with another, and can be used to discover how the values of one key variable can be used to indicate or predict the values of another variable. The analyst can thus identify leading indicators.

Other Features

Although SFII has no macros as such, any set of forecasts can be saved on a worksheet and to a file and thus serve as as a macro does to save later steps. SFII has no extensive internal printing utility, but all graphics can be printed by selecting GRAPHICS as a DOS command before starting and then using <CTl> and <Prt Sc>.

ANNEX A

Detailed Steps in READING Data into SFII Worktables.

```
SmartForecasts II     (C) 1985, 1986, 1987  Smart Software, Inc.

                    ............... ............... ...............
                        V1             V2             V3
                    ---------------------------------------------
............... C1 :          :              :              :
............... C2 :          :              :              :
............... C3 :          :              :              :
............... C4 :          :              :              :
............... C5 :          :              :              :
............... C6 :          :              :              :
............... C7 :          :              :              :
............... C8 :          :              :              :
............... C9 :          :              :              :
............... C10 :         :              :              :
............... C11 :         :              :              :
............... C12 :         :              :              :
                    ---------------------------------------------

EDIT MODE :

For help, type HELP EDIT or HELP followed by one of these commands:

CLEar  DEFine  DELete  EXPlore  FORecast  HELp  INSert   LABel  LISt
OOPs   QUIt    READ    REPlace  SAVe      SEE   SEQuence TITle  VARiables
```

SCREEN 1

User enters: READ, then ENTER

```
SmartForecasts II     (C) 1985, 1986, 1987  Smart Software, Inc.

                    ............... ............... ...............
                        V1             V2             V3
                    ---------------------------------------------
............... C1 :          :              :              :
............... C2 :          :              :              :
............... C3 :          :              :              :
............... C4 :          :              :              :
............... C5 :          :              :              :
............... C6 :          :              :              :
............... C7 :          :              :              :
............... C8 :          :              :              :
............... C9 :          :              :              :
............... C10 :         :              :              :
............... C11 :         :              :              :
............... C12 :         :              :              :
                    ---------------------------------------------

Do you still want to READ? (Press Y or N ):

WARNING

The READ command will replace the data table now on the screen.
```

SCREEN 2

User enters: Y

```
SmartForecasts II     (C) 1985, 1986, 1987  Smart Software, Inc.

                    ............... ............... ...............
                        V1             V2             V3
                    ---------------------------------------------
  ............ C1  :          :              :              :
............ C2  :          :              :              :
............ C3  :          :              :              :
............ C4  :          :              :              :
............ C5  :          :              :              :
............ C6  :          :              :              :
............ C7  :          :              :              :
............ C8  :          :              :              :
............ C9  :          :              :              :
............ C10 :          :              :              :
............ C11 :          :              :              :
............ C12 :          :              :              :
                    ---------------------------------------------

What do you want to READ? (Press W or L or A or D or T or Esc ):
Press  W  to read the  SFII WORKSHEET: ENROLLMENTS
Press  L  to read a    Lotus Wxx  data file.
Press  A  to read an   ASCII      data file.
Press  D  to read a    DIF        data file.
Press  T  to read the  TUTORIAL   data table.  To cancel, press  Esc .
```

SCREEN 3

Enter: A

```
SmartForecasts II     (C) 1985, 1986, 1987  Smart Software, Inc.

                    ............... ............... ...............
                        V1             V2             V3
                    ---------------------------------------------
............ C1  :          :              :              :
............ C2  :          :              :              :
............ C3  :          :              :              :
............ C4  :          :              :              :
............ C5  :          :              :              :
............ C6  :          :              :              :
............ C7  :          :              :              :
............ C8  :          :              :              :
............ C9  :          :              :              :
............ C10 :          :              :              :
............ C11 :          :              :              :
............ C12 :          :              :              :
                    ---------------------------------------------

Where is the ASCII file you want to READ? (Press A through Z or Esc ):

Press  A  or  B  for floppy disk drives  A:  or  B: .
Press  C  through  Z  for hard disk drives  C: through  Z: .

Press  Esc  to cancel the READ command and return to EDIT mode.
```

SCREEN 4

Enter: C

```
SmartForecasts II     (C) 1985, 1986, 1987  Smart Software, Inc.

                    ............  ............  ............
                       V1            V2            V3
                    ---------------------------------------------
............ C1  :        :             :             :
............ C2  :        :             :             :
............ C3  :        :             :             :
............ C4  :        :             :             :
............ C5  :        :             :             :
............ C6  :        :             :             :
............ C7  :        :             :             :
............ C8  :        :             :             :
............ C9  :        :             :             :
............ C10 :        :             :             :
............ C11 :        :             :             :
............ C12 :        :             :             :
                    ---------------------------------------------

Enter the path on C:
Enter  if the DOS volume has only a root directory
       or
Enter FULL path information to access any subdirectory (e.g.,\SALES\CARS ).

Press  Esc  to cancel the READ command and return to EDIT mode.
```

SCREEN 5

Enter: SF SFDATA then ENTE

```
SmartForecasts II     (C) 1985, 1986, 1987  Smart Software, Inc.

                    ............  ............  ............
                       V1            V2            V3
                    ---------------------------------------------
............ C1  :        :             :             :
............ C2  :        :             :             :
............ C3  :        :             :             :
............ C4  :        :             :             :
............ C5  :        :             :             :
............ C6  :        :             :             :
............ C7  :        :             :             :
............ C8  :        :             :             :
............ C9  :        :             :             :
............ C10 :        :             :             :
............ C11 :        :             :             :
............ C12 :        :             :             :
                    ---------------------------------------------

Enter the file name extension:

Enter the ASCII file name extension to be searched on
C:\SF\SFDATA
This extension must be 3 characters long, e.g.  TXT  or  PRN .
If your file name has no extension, press  Enter .  To cancel, press  Esc .
```

SCREEN 6

Enter: UNI then ENTER

The following ASCII UNI data files are stored on directory
C:\SF\SFDATA:

ARUGS	ARUGSG	ARUGSS	BAYES	CACDD	CAHDD	CASPC	CLASSNP
CFI	ENP	M1	NFDR	NNP	NNPC	NNPC1	NNPF
NNPX	BAVRAT	SSUN	TESTNP	TESTNPB	UN	UNCP	UNEM
UNEMEXF	UNEMP	UNEMPA	UNEMSSF	UNFD	XIP		

SCREEN 7

Enter M1 then ENTER

Enter a file name from the list above (or press Esc):

Press Esc to cancel the READ command and return to EDIT mode.

SCREEN 8

Does the ASCII file use the default format? (Press Y or N or Esc): Enter: Y

The default format has variable length field sizes, delimited by commas.
If the file uses the default format, press Y . If not, press N .
To cancel the READ command, press Esc .
Reading file C:\SF\SFDATA\M1.UNI

SCREEN 9

Do you want to READ the ASCII file by rows or columns? (Press R or C or Esc): Enter: C

Press R to read variables organized as rows into an SFII variable.
Press C to read variables organized as columns into an SFII variable.
Normally, you will press C to read a PRN file from Lotus 1-2-3 or Symphony.
To cancel the READ command, press Esc .

```
SmartForecasts II     (C) 1985, 1986, 1987  Smart Software, Inc.

                    ............... ............... ...............
                       V1             V2             V3
                    ---------------------------------------------
............... C1  :       112.030        :              :
............... C2  :       113.670        :              :
............... C3  :       114.930        :              :
............... C4  :       115.930        :              :
............... C5  :       117.130        :              :
............... C6  :       118.200        :              :
............... C7  :       119.700        :              :
............... C8  :       121.900        :              :
............... C9  :       123.500        :              :
............... C10 :       124.530        :              :
............... C11 :       125.800        :              :
............... C12 :       127.070        :              :
                    ---------------------------------------------

EDIT MODE :

For help, type  HELP EDIT  or  HELP  followed by one of these commands:

CLEar  DEFine  DELete  EXPlore  FORecast  HELp  INSert    LABel  LISt
DOPs   QUIt    READ    REPlace  SAVe      SEE   SEQuence  TITle  VARiables
```

SCREEN 10

Data now loaded in SFII Worktable

Enter: SEE C110

```
SmartForecasts II     (C) 1985, 1986, 1987  Smart Software, Inc.

                    ............... ............... ...............
                       V1             V2             V3
                    ---------------------------------------------
............... C105:       294.370        :              :
............... C106:       299.170        :              :
............... C107:       302.400        :              :
............... C108:       308.100        :              :
............... C109:       315.430        :              :
............... C110:       320.900        :              :
............... C111:       326.400        :              :
............... C112:       333.300        :              :
............... C113:       340.070        :              :
............... C114:       347.530        :              :
............... C115:       354.500        :              :
............... C116:       360.630        :              :
                    ---------------------------------------------

EDIT MODE :

For help, type  HELP EDIT  or  HELP  followed by one of these commands:

CLEar  DEFine  DELete  EXPlore  FORecast  HELp  INSert    LABel  LISt
DOPs   QUIt    READ    REPlace  SAVe      SEE   SEQuence  TITle  VARiables
```

SCREEN 11

Data scrolled to C110

Chapter 6

MANAGEMENT INFORMATION SYSTEMS

6.0 INTRODUCTION

The development of Management Information Systems (MIS) is a topic of great importance in educational planning and management. The effectiveness of the models and methods used by planners and policy makers (including those covered in earlier chapters) depends on the quality of information used. Information starts out as facts or data expressed in words or numbers. But, data only becomes information when it tells something useful. It is useful when it can be used in planning tasks.

What is a Management Information System? It is a structured collection of data that serves the main purposes of management: planning, control, and direction. This chapter will provide some guidelines for the design of Management Information Systems that can facilitate these purposes. But it is important to realize that the design of an MIS is not a routine or mechanical task. Every country and every system within a country is different. Thus, no simple set of procedures can assure a successful design. However, this chapter will present the major issues that a Ministry of Education should consider before moving into MIS.

There are three major phases in the design of a Management Information System:

1. Describe, analyze and chart the educational system, the organization and the main transactions in the organization.

2. Determine the educational requirements of the system.

3. Determine information requirements of the education system or organization.

What kind of planning and management tasks does the Ministry undertake? What kind of forecasting models and

Ministry undertake? What kind of forecasting models and methods are used? What kind of policy research needs to be conducted? The type and quantity of information required by the system for these and other important tasks should be determined at the start.

A. Identify sources of data to meet system information requirements

In the case of a Ministry of Education, the sources of data and information would include such things as national census, surveys, school census reports, manpower needs surveys, research reports, etc.

B. Enter data into computer databases (national, regional or local) or a network of smaller data bases in a form that promotes flexibility and easy access.

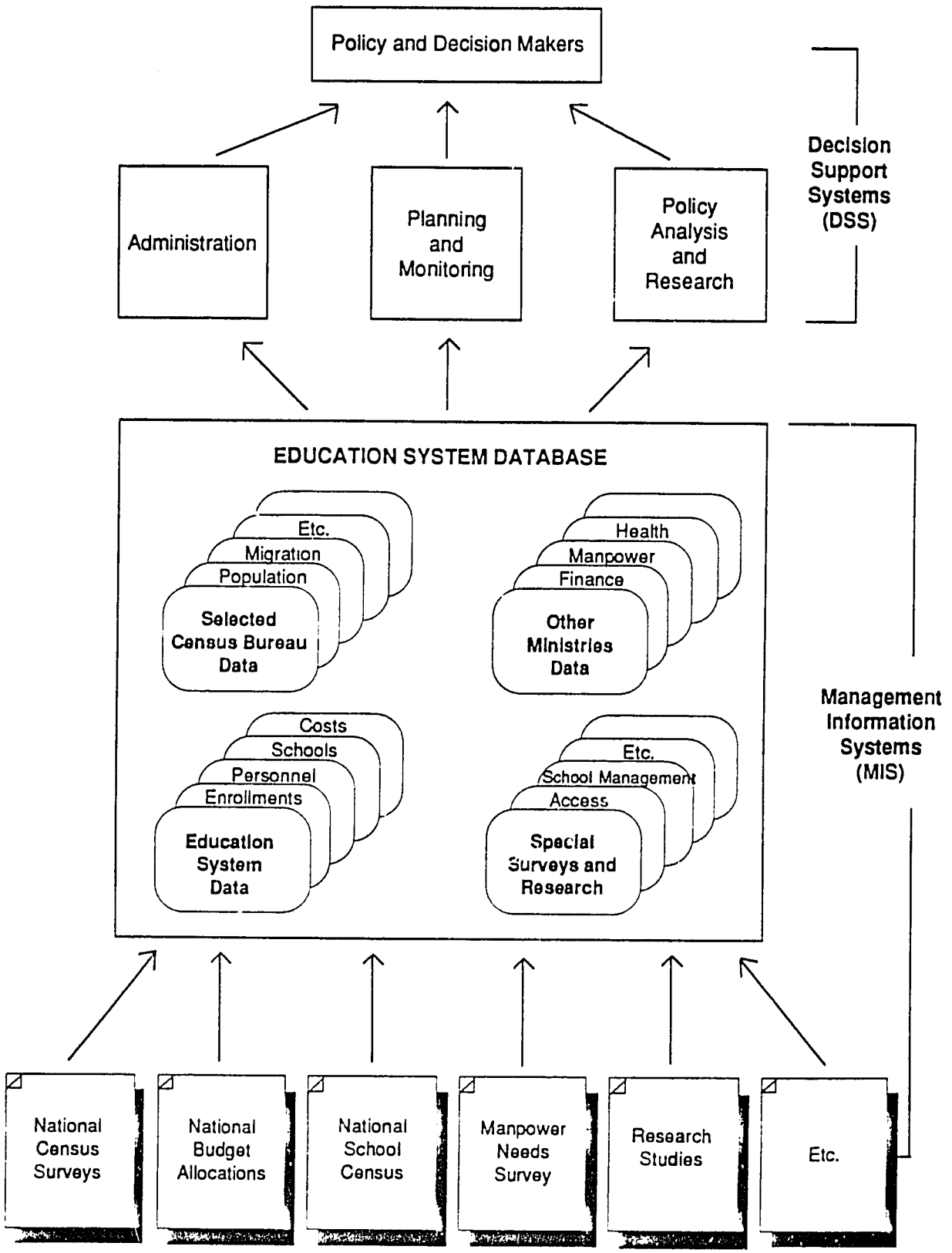
This procedure takes careful planning so that the data will be suitable for various types of analysis for present and future needs. (Software such as R:Base System V can be used to develop the structure of the database.)

Once the system databases are created and linked, officials can draw upon the databases when making decisions. When this occurs, the Management Information System becomes a Decision Support System (DSS). A Decision Support System is a Management Information System that is actually used as a framework for analysis and decision making. The development of Management Information Systems into Decision Support Systems is the whole point of work on information systems. The progression from data to information to management information systems to decision support systems is a long one. Yet, it is important to execute each step to insure that data and information that are collected are actually used in management and planning. (Use is far from assured, as surveys of practice demonstrate.)

Figure 6.1 on the next page is a diagram showing an ideal flow of information through an educational system centrally run in a Ministry of Education. The figure shows

[Note: Because of the similarity of the terms MIS and DSS, this chapter will primarily use MIS as an umbrella term. However, keep in mind that a Decision Support System is a Management Information System that has been developed to the point where it aids and improves policy, program and

Figure 6.1: Information Flow Diagram



data from many individual sources being combined in a central Education System Database. Information from the central database can then be used by officials in administration, planning and monitoring, and policy analysis and research who must make recommendations to the final decision makers.

6.1 BENEFITS OF MANAGEMENT INFORMATION SYSTEMS

The development of a management information system such as that shown on Figure 6.1 offers three main benefits to an educational system:

1. The data collected by different agencies and departments are stored in a central location or in national, regional or local bases that are linked. This makes information accessible to the decision makers in the Ministry of Education. Clearly, if the data are scattered in separated and unlinked, isolated locations, any attempt to draw upon them for research and policy making will be difficult (or impossible). A central database or a network of linked databases is required.

2. Data from different original sources can be pulled out and analyzed together for research and policy making purposes. This permits great flexibility in research projects because so many data elements are available in computerized format. For example, the School Census data category alone could include the following data elements:

A. School Identification
- Location
- Use
- Ownership and Responsibility

B. School Context
- Facilities
- Age
- Conditions

C. Students
- Enrollments by grade-sex-age
- Classification of students as new entrants, repeaters, and promotees
- Examination Results
- Number and Classification of Graduates

D. Personnel
- Administrators
- Teachers
- Other Professionals
- Support Staff
- Other

E. Costs
- Personnel
- Operational
- Capital

When entered into a management information system (MIS), any one of these elements can be (1) examined individually, (2) examined in relation to others from the School Census, or (3) examined in relation to data elements from any of the other instruments. Administrators, planners, and policy analysts can all take advantage of the central pool of data for different purposes:

A. MIS Uses in Administration

- Document preparation and disseminiation
- Budget Development
- Personnel and Student Record Keeping
- Presentations

B. MIS Uses in Planning and Monitoring

- Population forecasting
- Enrollment forecasting
- School mapping
- Scheme preparation and tracking
- Resource allocations, tracking, and projection
- Financing, costing, and cost projection
- Program and project planning

C. MIS Uses in Policy Analysis and Research

- Calculation and assessment of educational efficiency measures
    - Cohort analysis
    - Economic analysis
- Assessment of educational outcomes
    - Learning
    - Skilled manpower
- Assessment of equity in the system.

For any of these purposes, officials can draw information from the Education System Database and structure it for analysis using appropriate software packages:

1. Spreadsheets
2. Database management
3. Mapping programs
4. Statistical analysis software (e.g. SYSTAT)
5. Time series analysis and forecasting packages (e.g. SMARTForecasts II)
6. Comprehensive modeling systems (e.g. STEP)

For text handling and reporting after analysis, word processing and outliner packages are used.

3. Finally, a third benefit of management information systems (MIS) is that they can avoid duplication in data collection. The development of MIS calls for specifying what data are to be gathered and the form in which they are to be stored and accessed. This makes it clearer in advance which data are useful and used and which are not. A centrally designed database at the ministry level can reduce waste and duplication although the different departments and different regional entities may require slightly different information. As Figure 6.1 shows, the information to meet the various requirements can be stored in a general system and retrieved as needed for specific user purposes.

6.2 MANAGEMENT INFORMATION SYSTEMS: BACKGROUND INFORMATION

Alter (1976, 1980) lists seven main uses of computer based information systems:

1. Retrieval of isolated data items to answer specific queries. (Alter indicates this as the main use.)

2. Ad hoc analysis of data files

3. Pre-specified summaries of data in standard reports

4. Estimating consequences of proposed decisions

5. Proposing decisions

6. Making decisions

Although not specifically written for educators, it is clear that Alter's list of tasks closely parallels the data

1. Planners must often retrieve isolated facts to respond to specific questions of ministers and other high officials, or to inform users out in the provinces, districts or schools. There is much use of database information to answer ad hoc (special, one-time) queries. There would be more ad hoc use if ministries had their database in linked networks where it could be readily accessed for such use.

2. Sometimes the requests go beyond a few isolated facts and require more extensive analysis, study and reporting. Databases are used for research and special studies. Again this depends on the accessibility of the information.

3. Planners use data in forecasts in an attempt to anticipate future needs and problems.

4. Planners use data in spreadsheet models in the form of "what if" questions, or hypotheses, which can explore the consequences of alternative decisions. In this situation, the planners do not choose or make the decisions; rather, they present the options for choice and decision to higher officials. Spreadsheet models have made it easier for planners to:

a. Explore alternatives with "what if" hypotheses--"What would the total cost be if average senior teacher salary were $15,000?"

b. Test "sensitivity" by assessing percent change in one variable given percent changes in a related variable: "Percent change in unit operating costs for percent change in teaching costs."

5. Planners also estimate costs and assess the feasibility of new programs and projects that combine units of cost in new ways. The costs and benefits of the component elements may be known from experience and stored in the database. However, the exact combination in a proposed program or project may be new. Results and resource requirements can be tested against constraints, or limits on resources.

6. Finally, planners may also collect special data on programs and projects and use them to evaluate the effectiveness of the program tested and computed against personnel standards or norms. (This is the main purpose of the BRIDGES Project.) Planners may also compare programs for efficiency, where one program gives more result for the same cost than another (or gives the same result for less cost). Planners must evaluate whether the benefits exceed the costs and if so, in what ratio.

the benefits exceed the costs and if so, in what ratio. Most of these economic analyses and evaluation of educational programs depend on having a database with accessible and usable data.

Having the information needed by planners and policy makers in a computerized, linked and easily accessed databases will facilitate the accomplishment of the above tasks.

6.3 DECISION SUPPORT SYSTEMS: BACKGROUND INFORMATION

The concept of decision support systems (DSS) was developed to respond to criticisms of management information systems. Many writers said that MIS needed to be:

1. More adaptive,

2. Better able to deal with the problems of under specification that plague most education research,

3. More resilient, and thus better able to cope with situations of constant change,

4. More responsive and supportive of individual decision making, and

5. More socially sensitive to the need to develop organizational participation and consensus as a measure of effectiveness.

Some writers began to describe "decision support systems" that would overcome these failings. Keen (1980), for example, writes that a decision support system

1. Deals with semi-structured tasks,
2. Employs an evolutionary and "middle out" design strategy (i.e. a quick and dirty operations research model is introduced, tried, and later expanded or refined),
3. Draws on descriptive insights and normative models of decision making (heuristics, rules, learning, trial models used),
4. Intends to improve effectiveness rather than efficiency, and
5. Reflects an implementation strategy based on the use of skilled intermediaries, responsive service and programs that individuals find often today to interact with.

In another article, Keen (1981) enumerates the benefits of DSS. He say that decision support systems

1. Increase the number of alternatives examined,
2. Respond fast to the unexpected,
3. Facilitate ad hoc analysis,
4. Promote new insights and learning,
5. Improve communication, control, decision making at less cost and time, and
6. Promote more effective teamwork and better use of data resources.

In his later work, Alter (1980) also focused more on decision-making than simply on information management. Alter suggests this two-step process for the design of the system database in its final form:

1. Transform routine operations information into Management Information Systems (MIS), and then

2. Transform Management Information Systems into true Decision Support Structures (DSS).

Some lists of the virtues of DSS seem to be the positive side of all the older problems and issues for which management information systems were once criticized. However, writers find it easier to list the virtues of DSS than to give detailed steps for creating one.

The remainder of this chapter will present some general guidelines for the design of Management Information Systems. The topics that will be covered are: (1) Dealing with existing systems (2) MIS Design: The Team Approach (3) Determining Required Data and Information (4) Designing the MIS for Flexibility and Easy Access (5) Prototyping as a Development Strategy.

6.4 DEALING WITH EXISTING SYSTEMS

In an ideal situation, a system would begin the design of an MIS by determining exactly what data was desired and in what quantity and form it would be most useful. Then, instruments could be designed to collect the data exactly as desired. Finally, the data would be entered into the MIS in a way that promoted flexible and easy use.

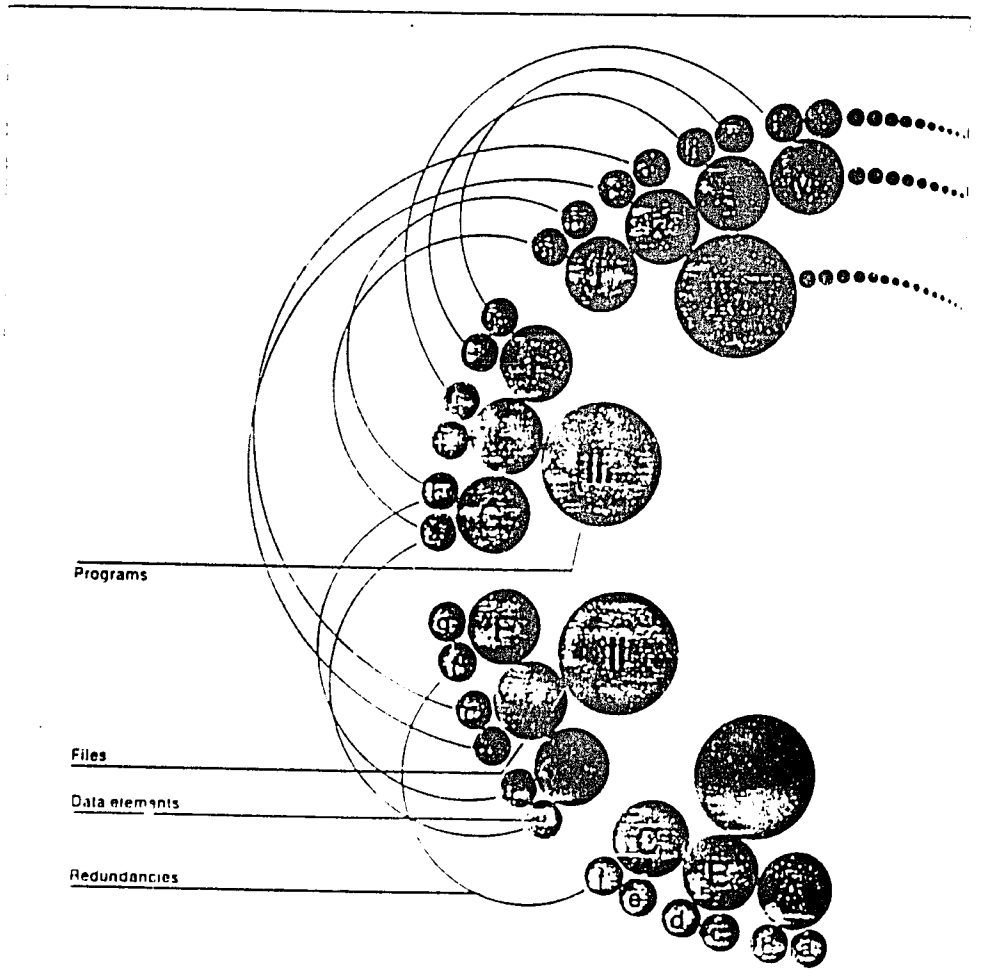
As usual, though, the ideal situation is rarely found. The reality is that education systems that attempt to move into MIS must deal with data collections (forms, procedures, tables, reports) and systems that already exist. Yet, the

tables, reports) and systems that already exist. Yet, the existing systems may not fit current information needs for management control or planning. Existing collections of school statistics have usually been built up over a long period of time. Often, the purpose or task they serve is unclear. Some data collection systems may have been started before planning even existed! Forms may have been borrowed from another place, where they served another purpose--or no purpose at all. Sometimes statistics did serve a purpose at one point, but that purpose no longer exists. Yet the collection of data sometimes goes on even after the purpose has been forgotten.

In general, older systems for dealing with data lack unity and flexibility. The data structures are not integrated. Rather, data of various types are simply stored in separate locations and it is impossible or difficult to analyze them together. These systems also tend to have a lot of redundancy (the same data is collected several times and stored in different files). Figure 6.2 below is an illustration of an older system.

Figure 6.2

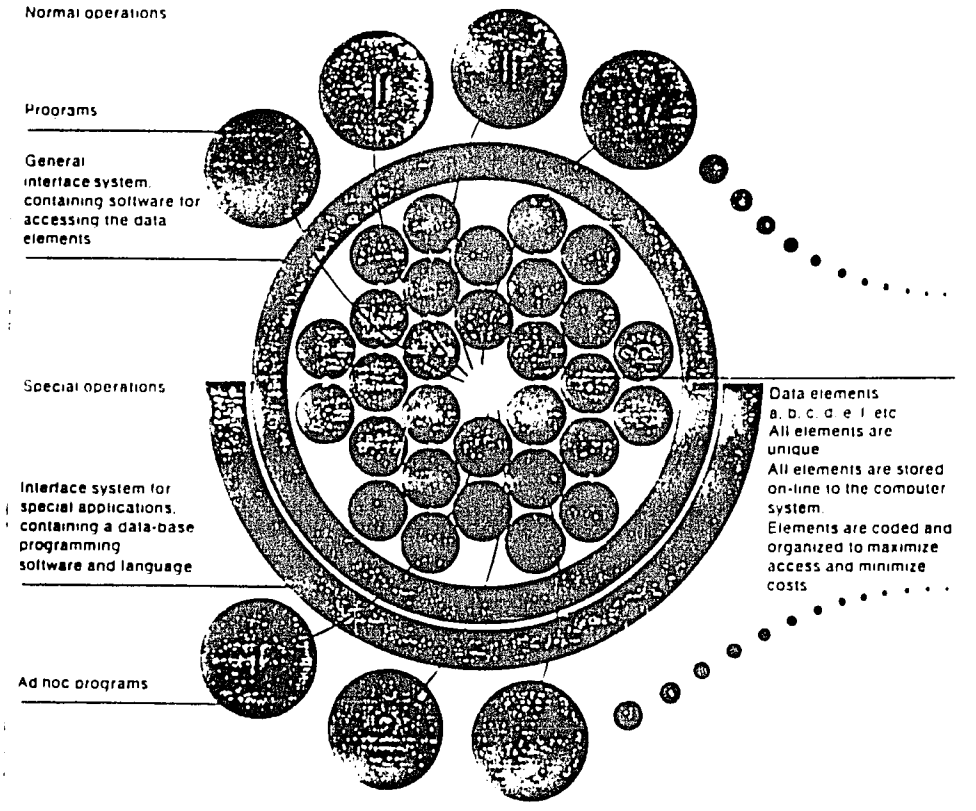
Older Information Database Structure
(Different, Separate Databases for Different Purposes)



On the other hand, newer computerized management information systems are integrated. They are organized so that all data goes into a common pool. This pool can be easily accessed within source supplier and user networks. Different data sets can be pulled out and analyzed together. Figure 6.3 below is an illustration of a newer information structure. It shows a definite contrast with Figure 6.2.

Figure 6.3

Newer Information Database Structure
(Pool of Data Elements; Accessed for Different Purposes)



[Source for Figure 6.2 and 6.3: Nolan, R.L. Harvard Business Review, Sept-Oct, 1973, p. 102-103]

Business Review, Sept-Oct, 1973, p. 102-103]

6.5 MIS DESIGN: THE TEAM APPROACH

The point of the previous section is that few (if any) systems begin with a blank page and design a new data system to fit their current tasks and purposes. Even MIS experts cannot simply appear with a new system ready to plug in. They too must deal with the older statistical collections that exist and work from there.

Given reality, then, how does a system begin the design of a Management Information System? The procedure that seems to work best is a team approach:

(1) MIS experts working with

(2) The people who produce the basic data and prepare the existing statistics, and

(3) The people who use the results.

This team must develop a new system of data forms, outcome indicators, and reports. Their work should build insofar as possible on elements of the old system. In no case should the old system be thrown away until a new system that is perceived as better is accepted for use. The two-fold process of (1) fitting the new MIS to the tasks (and the old system) and (2) developing use and acceptance by visiting experts, local experts, and end-users has a special name in MIS design. It is called prototyping. Prototyping means developing new information systems, trying them, adapting them and perfecting them. (More will be said about prototyping at a later point in the chapter.)

It is important to understand that a system is never "perfected". It is never perfected because it must serve current and future needs. Since future needs cannot be known, systems to meet these needs cannot be perfect.

Once the team has been established, there are two major tasks they must accomplish:

1. Data inputs required by the Education system and the sources of these data must be determined, and

2. The central database, and other databases to which it may be linked in anetwork, must be designed to be flexible and easily accessible to the users.

Each of these points will be discussed in more detail

6.6 DETERMINING REQUIRED DATA AND INFORMATION

An important part of MIS design is to assess existing sources of data and information. In national systems these include:

1. Regularly collected and reported data and information. In Egypt, for example, the major collection of information comes in the November census of schools. There are also collections shortly after the opening of the school year and in February. There are special forms and areas of coverage specified for these data. The November census covers schools (enrollments, facilities, etc.). The February census covers teachers and other personnel.

The data element on which most education operations data systems are built is the student unit, or "enrollment". Some form of pupil accounting system is used at all levels to track

a. student enrollment
b. enrollment flow
c. numbers of graduates
d. student programs and schedules
e. student loads and input ratios (pupil/teacher) used in estimating instructional input needs
f. student performance and test results
g. student exam results, marks/grades
h. administration
i. pupil personnel services
j. management personnel services
k. computerized libraries (and data bases)
l. education research.

2. Special surveys and test results collected for some special purpose. These are often collected from a sample designed to represent the values of variables in the entire system.

3. External censuses (the decennial population, housing, industry and agricultural censuses) and government and private industry reports.

4. Data generated as a by-product of on-going operations. For example, rosters of teaching personnel that are submitted for entry on the payroll. Or student data from attendance registers.

5. Data that is collected by other organizations for other purposes, but serves educational planners. For example, surveys of community attitudes, employment, private training needs, etc.

needs, etc.

Figure 6.4 below is an even a more complete list of the types of information that are useful to planners. Then, Figure 6.5 on the next page shows how the types of information listed below are used in various models and levels of planning.

Figure 6.5: Uses of Information in Educational Planning

PLANNING CONTEXT OR SETTING

Planning/Model of Method	Systems Level National	Systems Level State/Local	Institutional	Project/Program
Compound Systems Planning Models	All numbers 1 through 17			na
Resource Allocation Models for School Systems/Institutions	All numbers 1 through 17			na
Financial Planning Institutional	na	all 1-17		na
Costing/Budgeting Annual (Line Item) Systems/Institutions Projects	4, 6 8-11	4, 6 8-11	4, 6 8-11	na
Multi-Year PPBS Systems/Institutions	4, 6	4, 6	4, 6	na
Decision/Models (Choice)				na
Population Forecast System/Institution	1	1	1	
Enrollment Forecast, System/Institution	1, 6	1, 6	1, 6	
Cost Analysis System/Institution/Project	4, 6, 8-11	4, 6, 8-11	4, 6, 8-11	
Economic Growth and Manpower Demand	1, 2, 3, 6	1, 2, 3, 6	1, 2, 3, 6	
Social Demand (Cultural, Latent)	1, 5, 6	1, 5, 6	1, 5, 6	
Evaluation of Policy and Program/Project Alternatives	4, 6	4, 6	4, 6	4, 6
Policy Analysis System/Institution	4, 6, 8-15, 17	4, 6, 8-15, 17	4, 6, 8-15, 17	4, 6
Research/Development System/Institution/Project	4, 6, 16	4, 6, 16	4, 6, 16	4, 6, 16
Program Design Project Design	4, 6	4, 6	4, 6	4, 6

Note: Numbers in cells refer to numbers of information items in Figure 6.4.

Computer Based Systems of Educational Information:

In addition to the sources of information previously listed, there is also a vast array of computer-based sources of general and special information useful to educational planners. It is too large and varied to covered in detail. However, the major categories of computer-based information systems include:

1. Fact Data Bases cover a variety of information on almost all fields of human knowledge in almost all areas of the world;

2. Document Reference, Storage, and Retrieval Systems. The most comprehensive one in the field of education is ERIC, used by educators and planners in the US and throughout the world.

3. On-Line DataBase Systems. These entire systems of information can access many separate and individual data bases. They are available to most of the world through computer and phone link. INFACT, now the largest, provides computer access to over 650 individual databases,some of them with millions of records.

4. Search and Coordination Systems. So vast and complex are the data collections that special systems to find and use them are provided. DIALOG is one of the oldest and most used.

Once the team has decided what data and data sources should be included, the next step moves to matter of how much data is needed and in what format.

6.7 DESIGNING THE MIS FOR FLEXIBILITY AND EASY ACCESS

The previous section covered the range of different types of information that might become part of a management information system. Once that has been determined, what determines the quantity and quality of data that goes into a management information system? There are two key considerations:

1. The substance and quality of data entered into a MIS must be suitable for the tasks undertaken by planners and policy makers. Note that this should be determined before the data are collected.

2. The data must be collected and organized in a form and in an amount suited to the models and methods of planning and research. Here again, this is best considered before the data are collected.

Before any data is entered into a management information system, it should be determined that the data will

1. Fit the management and planning task or purpose for which it is intended,

2. Conform to the general standards of clarity, reliability (consistency), and validity (accuracy and relevance) that apply to data and information,

3. Fit the tasks or purposes that users of the existing databases and information collections perceive as their present needs. This means that the current users should be interviewed and asked to describe and define their present tasks and purposes.

4. Fit, to the extent possible, the future tasks and purposes of the system insofar as they can be anticipated. At least, the data should be adaptable to future tasks when they emerge (although this is difficult to assure).

In designing databases, the object is not to get perfection, but to insure adaptability. This is the feature of modern database management systems used for MIS. The data elements--the basic units of data--are stored for use in such a way that they can serve multiple purposes, present and future. The data are stored to answer multiple queries

(working questions) that users ask of the database. Newer database management packages like dBase or R:Base V store data for multiple use. The user designs his query to perform operations on the data (summing, averaging, joining, etc.) and output tables and reports according to need and use.

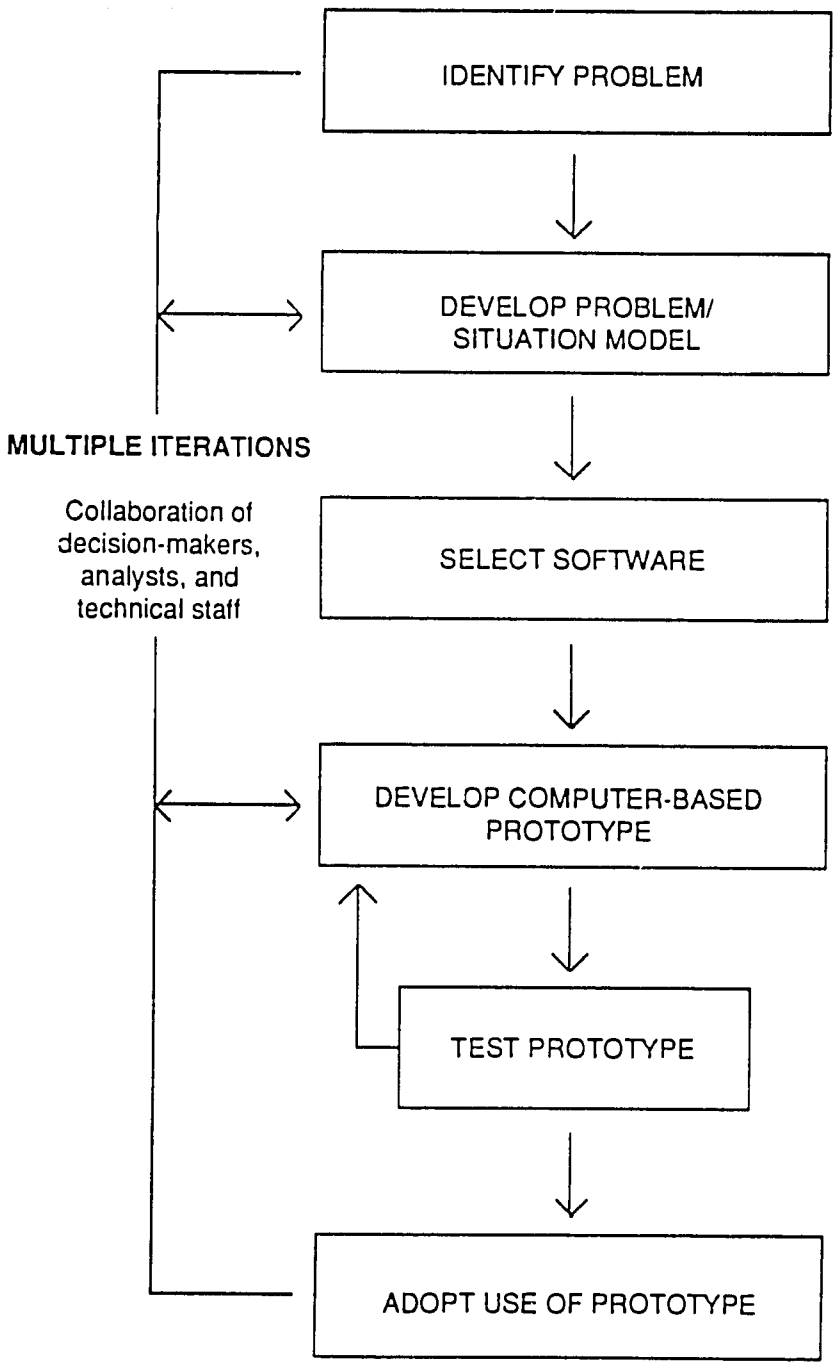
6.8 PROTOTYPING AS A DEVELOPMENT STRATEGY

How are the requirements of flexibility, individuality and systematization achieved in one and the same management information systems development project? Prototyping is one approach. With prototyping, analysts begin a process of development with a very simple representation of what the system will eventually be. At this early stage, the preferences of the intended users weigh as heavily as system technical requirements. After all, a system might be technically perfect, but be so difficult to use or lacking in essential elements that no one will ever use it. This problem can be avoided (or kept to a minimum) if all those who are expected to use the system have a chance to discuss the substance and form of the information before the system information codes are written in final form. Prototyping opens the system to discussion and shared decision-making. It is informational "glasnost"!

The prototyping process continues a gradual evolution through trial, test, critique, and modification of different versions. The users of the system should have hands-on involvement during this time. Their feedback will contribute greatly to the effectiveness of the ultimate system. Finally, the prototyping process will result with an implemented and functioning system. Because of the involvement of users throughout the development process, there should be no unpleasant surprises in the final stages.

Figure 6.6 on the next page shows a diagram of the prototyping process. Notice that the process calls for participation from the eventual users of the MIS as well as the technical staff.

Figure 6.6: The Prototyping Approach to Applications Development



A variety of tools have been developed to aid prototyping and to improve MIS development. Litaker, Staman and Rinker (1985) mention "fourth generation tools" as key to applications development. By fourth generation tools they mean many of the common computer programs and packages which are familiar to analysts and planners:

1. Spreadsheet models like 1-2-3 and IFPS. (Spreadsheet systems such as IFPS have built in interrogation commands to develop DSS ad hoc.)

2. Database and data management systems, starting with PC scale, relational bases like R Base V, dBase III+, running through the mid-sized hierarchical systems like FOCUS and RAMIS to large systems for mainframes like IDMS (Integrated Data Management System). (FOCUS and RAMIS are more than hierarchical database systems. They have a series of high level commands that query and develop prototypes and reports.)

3. New tools from artificial intelligence (AI) developments such as CLOUT and PROLOGUE. (CLOUT and PROLOGUE bring artificial intelligence to user system interface.)

4. Report generators including MARK IV and CULPRIT. (These packages allow query, sort, analysis and reports to be quickly developed.)

Specific to prototyping and MIS development design are a number of software tools for applications development. ADS/ONLINE (application development system), ACF (application control facility), and OLM (OnLine Mapping) enable systems developers to (1) generate ad hoc applications designs of both components and structures, and to (2) load data dictionary definitions, using an on-line dialogue generator. These tools execute the component tasks over the course of exploratory development, but also structure tasks interactively. Working directly with end-users, it is possible to develop a system that fits user needs--in iterative steps--from program design, to logic design, and only then on to program coding and testing.

New fourth generation programming languages can support developments by (1) opening it to end-use criticism from the outset, and by (2) maintaining systems order, flexibility, and openness at one and the same time. These tools serve analysts by reminding them that in the design and development of information systems for social planning:

> It is better to be <u>human</u> and pretend to be <u>rigorous</u> than to be <u>rigorous</u> and pretend to be <u>human</u>!

Exercise

RELATIONAL DATABASES

6.9 EXERCISE OBJECTIVES

The objectives of this exercise are:

1. To discuss the features of relational database packages and their uses for educational planning.

2. To review the basic skills needed to create a database (using LOTUS 1-2-3 for demonstration). This will include:

(a) How to create the structure of a database,

(b) How to sort, arrange, and analyze data,

(c) How to use database statistical functions.

6.10 INTRODUCTION TO RELATIONAL DATABASES

Today, relational database packages capable of maintaining millions of data records can be run on small computer (PC's) using software packages such as dBase or R:Base System V. The only limitation is the inconvenience of piling up too many floppy disks, or using up hard disk storage space to accomodate larger databases.

The basis of the relational base is a spreadsheet-like Row by Column matrix. This is obvious in Lotus 1-2-3 where a data base is written in same way as any spreadsheet. Once entered onto a spreadsheet database, the data fields and records can be manipulated (found, selected, compared, analyzed, summed, combined...) by a set of special Data command menus and data calculation functions.

The column names across the top of a spreadsheet database describe what the information in the columns below is about. Figure 6.7 on the next page shows a small example of data from 30 school buildings surveyed: The first column is the school name; The second column is the address; The third the number of classrooms, and so on. These are sometimes called the fields of information, but the term is not needed to understand a database. Relational databases can have hundreds of columns (R:Base V can have 800).

The rows of the database are sometimes called the records, but again the term is not important. In the example in Figure 6.7 each of the schools is a row or a record. In educational databases, common record units are students, teachers, schools, classrooms, and so on.

Figure 6.7

A Typical LOTUS 1-2-3 Database

				SCHOOL SURVEY DATA			($000)	(1=excel)
SCHNAME	ADDRESS	CRMS	LABWSH	LOT	HEAT	AGE	BOOK VALUE	CONDITION
Abbott	12 Bartholomew Sq	15	4	1.25	Solar	8	2900	1
Al'banha	46 Prospect Pl	12	2	0.90	Oil	22	900	2
Bogan	650 Rice Ave	12	2	1.00	Oil	25	920	3
Crane	903 Ray Rd	8	1	0.30	Oil	45	420	4
Denner	455 Daniels Rd	8	1	0.55	Elec	16	475	3
Dwane	12 Garden St	9	1	0.80	Elec	34	875	5
Francis	203 Somerset Ave	9	2	0.90	Gas	12	950	3

SCHNAME = School Name
ADDRESS = Address
CRMS = # of Classrooms
LABWSH = # of Laboratories and Workshops
LOT = Size of School Lot (Land Area)
AGE = Age of Facilities
BOOK VALUE = Recorded Value of Property

6.11 VARIETY OF DATABASE PACKAGES

All database packages have special features and all have good and bad points that fit some tasks well and other tasks not so well. Users can get help by reading reviews in computer journals and by working with the packages on actual tasks. It may take experience on several different applications to select the appropriate one. Only a few of the many possibilities are listed on the next page.

1. **Lotus 1-2-3**

LOTUS has a database management component that works as an integrated spreadsheet package. The strong feature is that the database feature and the spreadsheet analysis can be built into one model and application. The many logical, mathematical, and financial functions can be applied to the database by inserting a "D" into the usual function. For example, @DSUM( ) is the normal SUM function with a D added to it in order to add entries in a database. The same is true for other functions such as @AVERAGE, @DMAXIMUM, and others.

2. **R:Base System V**

R:Base (and to a lesser extent dBase) are examples of relational database packages that can handle a great deal of information. The data fields and records can be compared and combined and many different tables built. These packages can handle more data records more easily and more logically than LOTUS, but LOTUS offers some analytical functions R:Base does not have.

Both R:BASE and LOTUS can build special database models with macros in series of menus and commands. dBase has an even more powerful and versatile programming language for this. (However, dBase in more difficult to use than either LOTUS or R:Base.) LOTUS works best with structured information--i.e. when the records are standard in form and length. Other systems work better when the information comes in the form of text (words and sentences) and when the length of records varies.

3. **Notebook**

Notebook is a simple system that works well with textual information and when the records are of variable length.

4. PSF's and VP-Info

These packages are low in cost, but adequate for most tasks.

5. **RAMIS and FOCUS**

These packages run on large computers, e.g.IBM \370 CMS systems. They not only permit structuring databases, they also offer computer languages for developing systems models and templates for performing special planning tasks.

6. IDMS

This package runs network databases, but only on mainframe computers.

Having mentioned only some of the databases available, the remainder of this chapter will focus on the database capabilities of LOTUS 1-2-3.

6.12 LOTUS 1-2-3 DATABASE MANAGEMENT COMPONENT

The advantages of the LOTUS 1-2-3 database management component are:

1. It is integrated into a spreadsheet package and can use the general system commands and the mathematical and statistical functions to operate on data tables

2. It is simple to form and write the basic database on the worksheet using the columns as information fields, the rows as records, and the rows X columns to define a table.

The information which follows will give step-by-step directions for creating a LOTUS database.

A. Getting Started

(1) When the spreadsheet shows Ready, write the database on the spreadsheet. To do this, simply write the names of the columns onto the worksheet. On Figure 6.7, for example, the top row lists the fields, or names of the database columns. Each of the columns (SCHNAME, ADDRESS, etc.) is the name of a data field.

Each of the rows under each column is a record. In the case of Figure 6.7, the records listed under the field of "Schname" are the schools in the database: "Abbott" "Al'banha"; "Bogan"; "Crane", and so on.

(2) When you have finished writing the database, call the main menu using the / key. (Be sure the spreadsheet shows Ready before you do this.)

(3) Select the Data Command on the main menu and then choose the appropriate commands within the Data choices offered.

Figure 6.8 on the next page shows the LOTUS main menu with the Data menu on the second line.

Figure 6.8

LOTUS Main Menu and Data Menu

```
A1:                                                                    MENU
Worksheet Range Copy Move File Print Graph Data System Quit
Fill, Table, Sort, Query, Distribution, Matrix, Regression, Parse
```

	A	B	C	D	E	F	G	H
1					SCHOOL SURVEY DATA			($000)
2	SCHNAME	ADDRESS	CRMS	LABWSH	LOT	HEAT	AGE	BOOK VALUE
3								
4	Abbott	12 Bartholomew Sq	15	4	1.25	Solar	8	2900
5	Al'banha	46 Prospect Pl	12	2	0.90	Oil	22	900
6	Bogan	690 Rice Ave	12	2	1.00	Oil	25	920
7	Crane	903 Ray Rd	8	1	0.30	Oil	45	420
8	Demmer	455 Daniels Rd	8	1	0.55	Elec	16	475
9	Dwane	12 Garden St	9	1	0.90	Elec	34	875
10	Francis	203 Somerset Ave	9	2	0.90	Gas	12	950
11	Frome	34 Harley Pl	11	5	1.03	Solar	6	1800
12	Green	45 Lynwood Dr	7	2	0.60	Gas	45	550
13	Harriss	11 Pomona Rd	9	1	0.80	Oil	30	800
14	Inman	315 Fremont Ave	10	2	1.20	Elec	20	1600
15	Janson	95 Horne Pl	8	2	0.85	Gas	14	650
16	Kelly	19 Auburn St	8	2	1.00	Elec	12	1120
17	Mason	122 Stuyvesant Rd	11	2	1.25	Solar	7	1800
18	Nusbaum	87 Newbury St	10	2	1.00	Gas	17	1200
19	Opey	1122 Bellevue	10	3	1.30	Gas	9	1400
20	Peters	14 Brimshaw	12	2	1.20	Oil	32	1350

B. Using the Basic Data Commands

Select Data from the main menu (the first line in Figure 6.8). Within Data, there are eight choices:

- **Fill**: Fills a range of the worksheet with numbers
- **Table**: Creates a table of values (one or two-way classification)
- **Sort**: Sorts data records on primary or secondary key (basis on which the sorting is done)
- **Query**: Finds all data records satisfying given criteria, e.g. above a certain number; below the number; equal...
- **Distribution**: Calculates frequency distribution of numbers
- **Matrix**: Performs matrix math;e.g. pre multiplies or inverts.
- **Regression**: Calculates linear regression
- **Parse**: Handles arrangements of labels or words

C. Sorting and Arranging Data

1. Select Sort from the basic command Data.

2. Select Data Range to indicate the location of the database on the spreadsheet (include data entries only, not the field names).

3. Select Primary Key to sort (the column on which the sort will be run).

4. Indicate whether the data should be arranged in A (ascending) or D (descending) order.

5. Press Go to execute.

If two or more records have the same value in the first sort, the tie can be broken by selecting Secondary Key and running the sort again. The program sorts first on the primary key and then sorts the outcome of the first sort based on the secondary key that is identified, thus resolvuing the tie.

Reset removes directions out of database memory, so fresh directions can be given for a different sort.

D. Analyzing Data Relations: Query

To analyze data relations, select Query after the Data command. Figure 6.9 below shows the choices on the Query menu.

Figure 6.9

The Query Menu Line in Lotus

First line Data Menu choices:

Fill Table Sort Query Distribution Matrix Regression Parse

Second line Query choices:

Input Criterion Output Find Extract Unique Delete Reset Quit

For query, Lotus requires some arrangements to be made on the spreadsheet which tell the program (1) where to locate the fields and data inputs and (2) what criteria to apply in operating on the data or selecting fields

Data Range: Includes only the cells where the actual data are written and not the field names.

Input Range: Gives the cell locations of the entire database, including the field or column names row.

Criterion Range: The place on the spreadsheet where the criteria for selecting records are written. The Criterion Range is two or more rows of instruction written outside the database area. The first row specifies one or more field names; the second row gives criteria for matching records to fit. For example:

First Line: "CRMS" - This tells the program to select on the classroom column.
Second Line: ">6" - This tells the program to select schools with 6 or more classrooms.

The criteria can be written as:
(1) labels (a text entry in a column, such as to find schools classified as "Good"), or as
(2) numbers (a number value such as "6" to find schools with 6 classrooms), or formulas.

Formulas impose logical conditions for matching, e.g. telling the program to select records where cell values are greater, less, or equal to some criterion number or condition. Setting up a formula criteria works better if the analyst commands /Range Name Labels Down so that the program will run down through the records after testing the first.

Output Range: The place on the worksheet where the Query command results or records are extracted and written.

Figure 6.10 and Figure 6.11 on the next two pages show the data, input, criterion, and output ranges written on the earlier example.

Figure 6.10

The Query Menu Within Data

```
A1:                                                                    MENU
Worksheet  Range  Copy  Move  File  Print  Graph  Data  System  Quit
Fill, Table, Sort, Query, Distribution, Matrix, Regression, Parse
```

	A	B	C	D	E	F	G	H
1					SCHOOL SURVEY DATA			($000)
2	SCHNAME	ADDRESS	CRMS	LABWSH	LOT	HEAT	AGE	BOOK VALUE
3								
4	Abbott	12 Bartholomew Sq	15	4	1.25	Solar	8	2900
5	Albanna	46 Prospect Pl	12	2	0.90	Oil	22	900
6	Bogan	690 Rice Ave	12	2	1.00	Oil	25	920
7	Crane	903 Ray Rd	8	1	0.30	Oil	45	420
8	Demmer	455 Daniels Rd	8	1	0.55	Elec	16	475
9	Dwane	12 Garden St	9	1	0.80	Elec	34	875
10	Francis	203 Somerset Ave	9	2	0.90	Gas	12	950
11	Frome	34 Harley Pl	11	5	1.03	Solar	6	1800
12	Green	45 Lynwood Dr	7	2	0.60	Gas	45	550
13	Harriss	11 Pomona Rd	9	1	0.80	Oil	30	800
14	Inman	315 Fremont Ave	10	2	1.20	Elec	20	1600
15	Janson	95 Horne Pl	8	2	0.85	Gas	14	650
16	Kelly	19 Auburn St	8	2	1.00	Elec	12	1120
17	Mason	122 Stuyvesant Rd	11	3	1.25	Solar	7	1800
18	Nusbaum	87 Newbury St	10	2	1.00	Gas	17	1200
19	Opey	1122 Bellevue	10	3	1.00	Gas	9	1400
20	Peters	14 Grimshaw	12	2	1.20	Oil	32	1350
21	Pine	194 Maplewood	9	1	0.80	Oil	35	760
22	Pringle	19 Hill Rd	9	2	1.20	Oil	16	1100
23	Riley	1582 Central Ave	10	2	0.90	Elec	7	1250
24	Rothman	1 Pond Rd	15	5	2.00	Solar	6	2450
25	Staniy	384 High St	16	4	1.90	Solar	7	2650
26	Toner	2 Beech Ave	14	3	0.50	Gas	11	1750
27	Trish	131 Putnam Ave	12	2	0.75	Gas	12	1200
28	Udall	319 Sunnyside Dr	16	3	1.80	Solar	3	2850
29	Vance	16 Rice St	14	3	1.00	Elec	18	1800
30	Walton	86 Howard St	12	1	0.85	Oil	16	1050
31	Wilson	287 Carmel Terr	13	2	0.75	Solar	4	1940
32	Wright	1463 Woodlawn Ave	7	2	0.30	Elec	28	385
33	Yates	198 Upland Rd	10	2	0.85	Oil	27	1050
34	Zarod	714 Orchard St	12	3	1.50	Elec	15	1850
35								

```
36  Data Base Location Notes
37  Data Range A4...N34
38  Criterion Range D2...s3
39  Input Range A2...N34
40  Output Range D15...o100
04-Jan-90  03:04 AM
```

Figure 6.11

Criterion and Output Range Areas

```
N1:                                                                    MENU
Input  Criterion  Output  Find  Extract  Unique  Delete  Reset  Quit
Set the range containing data records
      N       O       P       Q       R       S       T       U
1
2             CRMS   LABWSH     AGE HEAT    BOOKVAL  CONDITION
3                0       0       1 solar        0        1
4
5
6
7
8
9          Sum
10           41795
11
12
13
14
15         SCHNAME     CRMS     AGE HEAT    BOOKVAL  CONDITION
16         Abbott        15       8 Solar     2900        1
17         Frome         11       6 Solar     1800        1
18         Stanly        16       7 Solar     2650        1
19
```

[Note that in the output range the column names must be written.]

E. Using the Query Commands: Find, Extract, Unique, Delete

Find: Find is the simplest Query command. After the Input Range and Criterion Range are indicated, the program locates and highlights those records that meet the criteria, beginning with the first record. The analyst can scroll down using direction arrows.

Extract: Extract requires more elaborate preparations on the worksheet. In addition to indicating the Input and Criterion ranges, the analyst must set up an Output area on the work sheet. Into this area, copy the names of the fields of interest and leave space for one or more records to be written that match the criterion. (Be careful not to run the output spill over any needed data on the worksheet.) When Extract is pressed, the records that meet the criterion match are copied into the output range area.

Unique: This command copies output and takes out duplicate records. There is also Delete, and Reset to start new search and copy sequences.

F. Using Database Statistical Functions

In addition to the commands just discussed, LOTUS 1-2-3 can apply the usual statistical functions Sum, Average, and so on by writing in a D to the usual function call. For example:

@DSUM(IR,3,CR)

The argument inside the parentheses "(IR,3,CR)" indicates the Input Range, Offset, and Criterion Range. Offset tells the column number to operate on. The first column of the database being 0; the 3 would mean column four.

G. Database Analytic Operations: Tables for Sensitivity

The LOTUS 1-2-3 DBMS component can do the analysis that spreadsheets do. It can "what if" database values or forms and recalculate the effects through the worksheet.

It can also do Sensitivity analysis by selecting Data and then Table. There are then two table options:

Table 1 tests the sensitivity of one or more output variables to changes in one input cell that is made to take on varying values.

Table 2 tests the sensitivity to variations in two different input cells taking on varying values. In both Table 1 and Table 2, the analyst sets up a sensitivity table area off on the side of the worksheet and follows this procedure:

1. Type a column of varying values for the input variables.

2. Type a row of cell references or formulas for the output variables. Leave the remaining rows of the table blank, to be filled in with the sensitivity analysis. The area or Table Range includes

   (a) a column for the input value that is being varied, and
   (b) columns for the output variables that are being tested for sensitivity.

(Note: The Table Range area should also include a row at the top which references the cell formulas where the variables of interest are. The Input Cell is then entered and the sensitivity analysis results are filled into the empty rows of the table in the table range area.)

H. More Advanced Data Analysis Operations

There are three more choices on the Data menu (shown on Figure 6.8) that have not yet been described. They include some very useful data analysis techniques:

Distribution: This command (1) creates a frequency distribution out of a set of numbers (i.e. arranges the number values into groups or intervals) and (2) gives the frequency (number of occurrences) of the values within each interval. For example, the basic population by age groups is a frequency distribution; within each age group (0-4, 5-9, 10-14, etc.), the number that fall within that group would by given as a result of the Distribution command.
(Note: Bin Range is the place on the spreadsheet where you tell the program the interval values you wish to use.)

Matrix: This command will invert a matrix of numbers (arrangements of numbers by rows and columns as in the basic enrollment model which shows a matrix of flow coefficients). It will invert a square matrix up to a size of 90 rows and 90 columns. The matrix is inverted in order to solve a set of simultaneous equations. For example, in the flow model

$$A_{(t)} X_{(t)} = X_{(t+1)}$$

the inverse of matrix A (written A^{-}) is the matrix which when multiplied by A will give I or the identity matrix. It results from applying transformations that reduce the coefficients of the variables in a set of equations to 1. It solves the equation set.

Regression: This command will yield a coefficient that shows how much one variable (the dependent) changes as one or more other variables (independent) change. It shows dependency relations that are useful in understanding variables in planning analysis. The general operations of the database management program in LOTUS can be used without knowing these statistical analyses commands, but they are useful and worth knowing and applying.

APPENDIX A

DATABASE FORMULAR LISTINGS

```
O2: "CRMS
P2: "LABWSH
Q2: "AGE
R2: 'HEAT
S2: 'BOOKVAL
T2: 'CONDITION
O3: +C3>10
P3: +D3>3
Q3: +G3<12
R3: 'solar
S3: +H3>1000
T3: +I3<2
O9: 'Sum
O10: @DSUM(A2..N34,7,O9..P10)
O15: 'SCHNAME
P15: "CRMS
Q15: "AGE
R15: (F0) 'HEAT
S15: 'BOOKVAL
T15: 'CONDITION
O16: 'Abbott
P16: (F0) 15
Q16: (F0) 8
R16: 'Solar
S16: (F0) 2900
T16: (F0) 1
O17: 'Frome
P17: (F0) 11
Q17: (F0) 6
R17: 'Solar
S17: (F0) 1800
T17: (F0) 1
O18: 'Stanly
P18: (F0) 16
Q18: (F0) 7
R18: 'Solar
S18: (F0) 2650
T18: (F0) 1
```

ASSIGNMENT #1: EVALUATION OF AN EDUCATIONAL PLAN

PURPOSE: The purpose of the exercise is for you to become familiar with educational plans from countries other than your own. This will allow you to analyze the similarities and differences in how countries approach educational planning. You may select a national plan, an institutional plan, or a project plan. You can then consider how the plan from your own country relates to others. The exercise will also be a review of the material on goals and objectives presented in class and in Chapter 1 of the text.

DIRECTIONS: A selection of plans is on reserve in Gutman Library. Select one from a country other than your own. Before reading the plan, read the evaluation criteria that begins on the next page. Then, as you study the plan, fill Part One of the evaluation form. Most items can be answered with a check mark or circled. However, feel free to write additional comments. Leave blank any items that do not apply. For Part Two of the form, give your summary opinion of the plan.

This assignment should be completed individually. You will turn in the entire evaluation form on Thursday morning, August 4.

REFERENCE: "Educational Planning," Chapter 1.

Title of Selected Plan _____

Country _____

PART ONE: EVALUATION CRITERIA

I. What type of plan is it? (Check the correct answer.)

__ Systems Plan (National school system, regional system or sub-system; i.e., linked institutions?)

__ Institutional Plan (Note: Large institutions may also be a type of system if institutional sub-units are viewed as sub-systems, but usually a plan for a single institution, even a university, is different from a system plan.)

__ Project Plan (Usually a project has a limited life and clear, specific objectives.)

__ Program Plan (The distinction between project and program is not always clear. Usually a project is free-standing and may exist outside of an institutional setting, and a program may be a sub-part of an institutional mission.)

__ Other (Some plans cut across types. For example, the support of a particular program activity; e.g., vocational education programs for handicapped, which may be carried out in several institutions not linked in the system.

Describe: _____

II. What is the scope of the plan?

__ General society, economy (political and cultural contexts described?)

__ Educational sector only (relation to other sectors described?)

__ Total school system (all major levels)

__ Complete level (Primary, Secondary, Higher Ed.)

__ Institution

__ Sub-Institutional (school division or department)

__ Other. Describe _____

III. What is the level of diagnosis and assessment?

A. Is the past performance assessed? YES ____ NO ____

B. Is the present situation diagnosed and analyzed? YES ___ NO ___

If yes, which of these components are addressed?

__ Activity levels (e.g., enrollments)
__ Resources (e.g., teachers)
__ Organizational capacity

C. Are current policies reviewed and consequences briefly assessed? YES ___ NO ___

IV. What is the domain of the plan; i.e., what is the geographical or political area covered?

__ National
__ Regional, state province
__ District
__ Municipal, local

V. What is the time horizon for the plan? (Check one.)

__ Long-range (ten years or more)

__ Mid-range (more than one year, usually three to five)

__ Annual Operating Plan (If so, is it related both to the annual budget and multi-year plan goals? YES ___ NO ___

__ Very Long Range (If so, is it a real plan perspective or merely futurology exercise? Describe: _____

VI. What is the authority and power structure behind the plan? Describe: _____

_____

Is the plan a mere compliance document; i.e., designed to show that activities mandated by legislation are being carried out? YES ___ NO ____

Does the plan take into account all the relevant agencies which might hurt or help the plan and whose interests are affected in some way by the plan and the policies it implies? YES ___ NO ___

VII. Are goals, missions, objectives, and targets clearly specified and put into a reasonable hierarchy?

YES ___ NO ___

If so, describe the hierarchical structure and organization of the plan: _____

_____

_____

Is there confusion by lack of hierarchical definition, such as when specific activities are referred to as goals, objectives, or missions, etc.?
YES ___ NO ___

Does the plan describe to what extent earlier objectives were attained? YES ____ NO ____

Does it indicate to what extent (and which) agencies will contribute to the accomplishment of future objectives? YES ___ NO ____

VIII. Is there a plan to evaluate the outcome of the plan?

YES ___ NO ___

Are there indicators that will show attainment of goals or objectives under the plan? YES ___ NO ___

Are there systematic tests of the outputs and objectives? YES ___ NO ___

IX. Is there a systems chart or model of the school or training system covered by the plan? YES ___ NO ___

Is the chart complete and comprehensive? YES ___ NO ___
Is it clear? YES ___ NO ___

Does the chart link the major sub-systems or components? YES ___ NO ___

X. Is there a model of the organization (an organizational chart) that will carry out the plan?
YES ___ NO ____

PART TWO: SUMMARY REVIEW

In your own words, write a statement of your opinion of the plan. Emphasize what you feel to be the strengths and weaknesses of the plan. You may use additional sheets of paper if you desire.

Your name _____

PART TWO: SUMMARY REVIEW

In your own words, write a statement of your opinion of the plan. Emphasize what you feel to be the strengths and weaknesses of the plan. You may use additional sheets of paper if you desire.

Your name _____

ASSIGNMENT #2: EVALUATION OF A SOFTWARE PACKAGE OR MODEL.

PURPOSE: The purpose of the exercise is to provide you with an opportunity to examine and critically analyze a computer template model or a computer package. A "package" is a product (usually copyrighted) released by a manufacturer; a "template" or "model" is a computer representation of some reality intended to help planners understand specific problems or issues better.

Your task is to (1) explore a model or package to determine its quality and its usefulness as a planning tool and to (2) write an effective and useful review so that other planners can benefit from your comments.

Another purpose of the assignment is to present a comprehensive method of evaluating software. You may want to adapt the software evaluation form that follows for your own Ministry or department after returning home.

DIRECTIONS: Select a package or model from among those listed on the next page. (With the instructor's approval, you may review one not on this list--as long as it is relevant to educational planning.)

Use the evaluation form which follows to help structure your review. Complete any items that apply to the software you are reviewing.

MODELS

EDFISMO--Education Finance Simulation Model by Zymelman. (Projects student flow, teacher requirements, costs, and budget requirements.)

EDSIM--Education and Demographic Simulation Model.

EDP MENU--(from World Bank Education Program).

ECC--Economies of Curricular Choice (Lotus template model).

POPEX2--Lotus template model by Davis/Cassidy.

SCHIEFELBEIN/CUADRA--BRIDGES Lotus model.

NCHEMS--A series of "Decision Support" LOTUS templates from the National Center for Higher Education Management Studies. Templates include: POPULATN, ENROLLEE, STUFLOW, FACULTY, FINOVIEW, FINPLAN, FINTRADE.

Computer Flow Model for Pupil's Analysis (Nassir Amin, AEPAM, Pakistan--Lotus template)

Education Projection and Simulation Model (Mohammed Ikram Qureshi, AEPAN, Pakistan)

STEP--Package Developed by RTI for BRIDGES.

PACKAGES

LOTUS 123--Package from Lotus Development Corp.

MULTIPLAN-Spreadsheet Package from Microsoft Corp.

JAVELIN--Spreadsheet package.

IPFS/PERSONAL--Interactive Financial Planning System (for educational planners).

STELLA--(runs on McIntosh).

SUPER PROJECT PLUS--PERT/GANNT project planning and management software.

NOTEBOOK and NUTSHELL--Simple, inexpensive database generators.

SOFTWARE EVALUATION FORM

General Information

NAME OF PACKAGE OR MODEL:

_____

OWNERSHIP: (Company or person owning the rights to the package or model) _____

PRICE: _____

COPY STATUS:
Is the software copyrighted? YES ____ NO ____

Copy protected? YES ____ NO ____ If so, specify the details (i.e., # of copies permissible, can you copy down to hard disk?) Describe: _____

_____

VERSION(S)/RELEASE DATES:
What is the version of the package reviewed? _____

How long has the current version been available in the field? _____

If a model, when and where was it developed?

_____

AVAILABILITY: (How easy or difficult is it to obtain the software?) _____

Give specific source where the software can be obtained, if appropriate. _____

_____

DOCUMENTATION: (Is documentation included with the software?) YES ___ NO ___. Is the documentation easy to read, well organized, and helpful? YES ___ NO ___.

TUTORIALS: (Are any tutorials included to help the user learn the program? YES ___ NO ___.

REVIEWS: (List any books, journals, newspapers, etc. in which the package was reviewed and give dates.)

_____

_____

What other products was the package compared with?

_____

FIELDTESTING: (If the information is available, where and when was the program fieldtested in an educational setting?) _____

Who evaluated it? _____.

Can other users of the software be contacted?

YES ___ NO ____

Is there a users' group which provides mutual support?

YES ___ NO ___

Technical Information

HARDWARE REQUIREMENTS:

IBM (compatible) ___ Apple ___ Other _____

Amount of memory required _____

Is a hard disk or co-processor required? YES __ NO __

Is companion software required or desirable? _____

SYSTEM/PACKAGE: (If a model, what programming language or software package was it constructed with?)

_____

Can the data format be converted to match other popular packages? ASCII ___ LOTUS ___ R:BASE ___
Other _____

MENUS: Is the software "menu-driven"? YES __ NO ___

If not, would menus help, or just slow it down? _____

COMMAND LANGUAGE: (How easy or difficult is the programming or command language that must be used to operate the program?_____

_____

Is there "on-line" help for the user? YES ___ NO ___

EASE OF USE: (How hard or difficult is it to: Enter, edit, and delete data? Create new applications? Create reports?)_____

_____

GRAPHING: (Is graphing of output possible? How hard or easy is it to do?) _____

_____

Is the output clearly understandable? YES ___ NO ___

SAFEGUARDS: (What are the safeguards to protect against inadvertent destruction or modification of the application or the data?) _____

_____

Are there automatic "re-build" features (to re-build a file if it becomes corrupted?) YES __ NO ___

How does the package handle errors; i.e., what kind of error checking and reporting capabilities does it have?

_____

RELIABILITY: (How reliable is the package/model; i.e, does it ever accidentally exit to the operating system?) YES ___ NO ___

Do the error messages help the user avoid problems? YES ___ NO ___

What is your overall reaction to the technical quality of the software? EXCELLENT ___ GOOD ___ NOT GOOD ___

Would the program be easy to demonstrate to non-technicians? YES ___ NO ___

Application Information

PURPOSE & STRUCTURE: (What is the general purpose of the model or package--what problems does it address?)

_____

_____

What general approach guides the construction of the software? _____

_____

If a model, is there clear specification of assumptions underlying and supporting the model? YES ___ NO ____

DATA REQUIREMENTS: (If a model, what data is required by the model--if any?) _____

_____

PLANNING TASK USES: (What specific planning-related tasks is the model or package designed to accomplish?)

_____

Examples:
-- Problem conceptualization (e.g. heuristics)
-- Systems conceptualization
-- Policy research and evaluation (quant. or qual.)
-- Project/program design (sched, costing, work flows)
-- Institutional program design
-- Operations management and control
  -- Costing and Budgeting
  -- Resource scheduling
-- Database
-- Strategic Planning:
  -- Forecasting--trend analysis,curve fitting, etc.
  -- Projections--economic, population, enrollment resource needs.
  -- Resource allocation (comparison of alternative scenarios).

PLAN DOMAIN USES: (For what level was the model or package developed--systems, institutional, program, project, etc.?) _____

DECISION SUPPORT FEATURES: (Does the software allow for any of the following: Sensitivity analysis ("What if"); Goal Seeking; Impact analysis; Comparison of alternative scenarios?) YES ___ NO ___ Describe: ___

_____

Evaluation Summary

Does the software accomplish what it sets out to? In other words, how valid is it? How effectively or efficiently does it accomplish its objectives?

What are the limitations of the approach the software tak for achieving the the stated purposes (if any)?

What do you feel to be the major strengths of the softwar

What do your feel to be the major weaknesses of the software?

How do you believe the software could be improved?

On a scale of 1 to 10 (ten is highest or best), how would you rate the software in relation to its relevance and usefulness to

Resolving major problems of planning:

1 2 3 4 5 6 7 8 9 10 (Circle)

Serving real world needs:

1 2 3 4 5 6 7 8 9 10

After completing this review, would you purchase this software or recommend it to anyone else?
YES ___ NO ___

Please make any additional comments that would help someone in deciding whether or not to buy this package.

Specify the tasks and operations that you or other reviewers performed in reviewing the package.

Examples:

--Creating a new application (formatting and defining the application, specifying relationships and/or writing formulas.
--Entering, editing and deleting data and/or existing application structures.
--Creating and printing reports and graphs.
--Importing/exporting data with other packages, determining the difficulty of transferability.

REVIEWED BY: _____ DATE: _____

_____

ASSIGNMENT #3: ENROLLMENT FLOW ANALYSIS EXERCISE

PURPOSE: The purpose of the exercise is to develop an enrollment model for doing diagonal analysis with an "electronic spreadsheet" computer-software package. You will become familiar with Lotus 1-2-3 including new features, such as matrix multiplication. The model you will build simulates the numbers of students who pass through a graded school system. The computer model will be designed on the basis of the assumptions that underlie an enrollment flow model.

DIRECTIONS: This exercise should be completed with two people working together. You can do the entire exercise by setting up only three tables on the same spreadsheet. (HINT: Use the outputs of one table to feed the inputs of the next table and use the copy command as often as possible to copy formulas which are the same in successive years. Thus, you will save a lot of retyping.) More detailed instructions are on the next page.

REFERENCES: "Educational Planning," Chapter 4.

MORE DIRECTIONS:

1. Using LOTUS 1-2-3, develop a diagonal analysis model to get one set of transistion rates (p's, r's, and d's) for grades 1 through 6 for 1971 to 1976. Use the following tables to help you:

Table C.1 has the enrollment data for the assignment.

Table C.2 has information on repeaters.

Table C.3 and and information on page 5 of this assignment will help you develop your diagonal model.

2. After computing the flow rates, arrange the p's, r's, and d's in a matrix similar to Table 4.6 in Chapter 4 of "Educational Planning".

3. Then use the flow matrix you develop in 2 above to forecast enrollments from 1977 to 1980. You may use the average of the five year's rates. The number of new entrants to Grade 1 are projected to be:

1977	103,217
1978	106,314
1979	109,502
1980	112,788

You may wish to create a summary analysis of your results similar to Figure 4.2 on page 21 in "Educational Planning", Chapter 4.

Table C.1

Enrollments by Grades, 1964-1976

Year	Totals	1st	2nd	3rd	4th	5th	6th
1964	344637	117344	81181	51805	41077	27828	19102
1965	356998	118822	84215	59867	43054	29727	21313
1966	365932	119531	84921	62729	44796	31438	22517
1967	378463	120810	88608	64441	47409	32968	24227
1968	399685	120544	94760	69050	51103	36859	27369
1969	408524	119784	94320	72769	53250	38819	29582
1970	424179	125052	96799	74661	56602	40725	30340
1971	436857	126034	99585	77467	58009	43397	32365
1972	440151	125983	98826	79252	59114	43878	33098
1973	451530	126077	100796	80675	62147	46789	35046
1974	454853	124709	100324	82160	63299	47874	36487
1975	452249	121736	99139	81889	63763	48283	37439
1976	467552	128537	100560	83343	65693	50249	39170

Source: Unpublished figures, Dirección de Planeamiento, MEC.

Table C.2

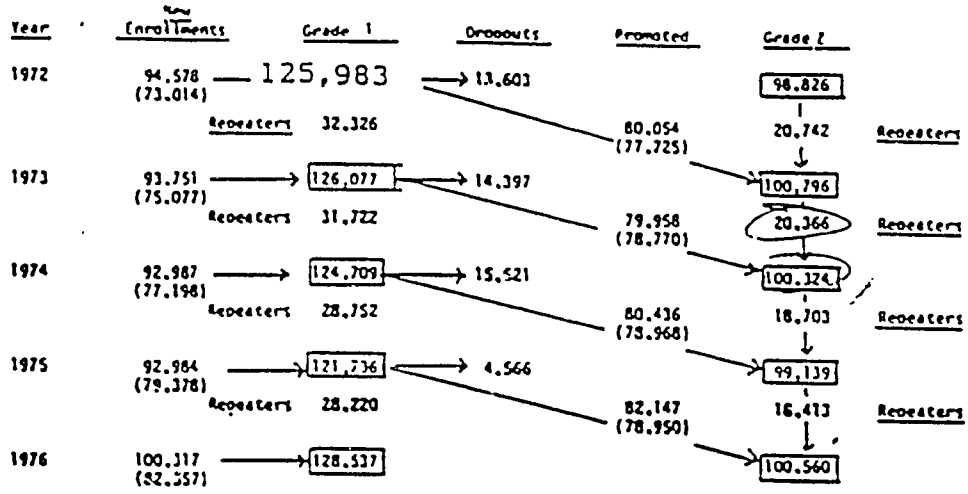
Repeaters and Repetition Rates by Grades: 1964-1976

Year	Total	Grades											
		1st	r	2nd	r	3rd	r	4th	r	5th	r	6th	r
1964	48274	32652		18397		11051		6074		2894		1385	
1965	78165	34679		20360		11331		6673		3230		1292	
1966	82486	36216		21262		12773		7391		3393		1451	
1967	85920	36518		22581		13805		7526		3728		1762	
1968	67623	31765		18339		9524		4812		1994		1189	
1969	73251	32082		20506		11044		5794		2525		1300	
1970	74943	31501		21096		12183		6067		2803		1293	
1971	76449	32265		21024		12389		6320		2960		1491	
1972	77169	32021		20890		12583		6909		3288		1478	
1973	77642	32326		20742		12161		6659		3330		1424	
1974	76366	31722		20366		12856		6963		3183		1276	
1975	69378	28752		18703		11946		6103		2680		1186	
1976	68606	28220		18413		11814		6155		2754		1250	

Source: Unpublished figures, Dirección de Planeamiento

Table C.3

Classic Method of Diagonal Analysis



Help For Assignment #3

(1)	a. Students in 2nd grade in 1974 (Table C.1)	100,324	
	b. Repeaters in 2nd grade in 1974 (Table C.2)	20,366	
	$(a - b = c)$ c		
	c. *Promoted* from 1st grade in 1973 (computed)	79,958	
(2)	d. Students in 1st grade in 1973 (Table C.1)		126,077
	e. Repeaters in 1st grade in 1974 (Table C.2)	31,722	
	c. Promoted to 2nd grade in 1973 (computed above)	79,958	
	$(e + c = f)$		
	f. Repeaters plus promoted (computed)		111,680
	g. Drop-outs from 1st grade in 1973 (computed $d - f = g$)		14,397
(3)	h. Students in 1st grade in 1974 (Table C.1)	124,709	
	e. Repeaters in 1st grade in 1974 (Table C.2)	31,722	
	i. *Newcomers* in 1st grade in 1974 (computed $h - e = i$)	92,987	

The same procedure is repeated in the next year:

(1)	a' Students in 2nd grade in 1975 (Table C.1)	99,139	
	b' Repeaters in 2nd grade in 1975 (Table C.2)	18,703	
	c' *Promoted* from 1st grade in 1974 (computed $a' - b' = c'$)	80,436	
(2)	d' Students in 1st grade in 1974 (Table C1)		124,709
	e' Repeaters in 1st grade in 1975 (Table C2)	28,752	
	c' Promoted to 2nd grade in 1974 (computed above)	80,436	
	$(e' + c' = f')$		
	f' Repeaters plus promoted (computed)		109,188
	g' Drop-outs from 1st grade in 1974 (computed $d' - f' = g'$)		15,521

HINTS ON MATRIX MULTIPLICATION WITH "Lotus 1-2-3"

Multiplying matrices is made simple with Lotus. Keep in mind that when multiplying matrices, however, there must be the same number of columns in the first range of numbers as there are rows in the second range of numbers.

You will do well to practice setting up a matrix and experimenting with it, in order to see what is going on. For example, we suggest you try the following set up and run through the following multiplication. Then experiment with some numbers of your own. (Unlike the example, BE SURE NOT TO LEAVE ANY SPACES BETWEEN THE ROWS.)

First set up the following numbers in the indicated cell positions on a new Lotus worksheet. We are leaving the first row (row #1) and column D blank just to allow some space for visual clarity. (NOTE: These numbers will not be surrounded by boxes on the Lotus screen.)

Columns

Rows	A	B	C	D	E	F
1						
2	3	2	1		2	
3	4	3	2		5	
4	5	4	3		4	

Now pull up the menu bar by pressing the "/" key and then the ENTER key.

Press "D" for Data, press ENTER,
Press "M" for Matrix, press ENTER, and then
Press "M" again (or move the cursor to the right) for MULTIPLY; Press ENTER.

You will then be asked to specify the first range to multiply. Position the cursor (if you have not already) on the cell which contains the number "3" (that will be cell position A2). The menu bar at the top of the screen should now show you that the number 3 is in cell A2.

Now press the PERIOD KEY and two dots (or periods) will appear to the right of the column/row designation (A2) at the top of the screen. Now the range is anchored. Next move the CURSOR-RIGHT key (the arrow under the number 6) so that the cell position indicator is two cells to the right and two cells down. Now you should see all nine numbers (all the numbers in our left-most group of numbers) high-lighted by the cursor. The cell position at the top of your screen should now look like this:

A2..C4

Press ENTER and you have specified the first range.

Next, the computer will ask you to specify or enter the range address or range name of the second matrix (the one you wish to multiply times the first). Now do the same thing as you did before: Move the cursor to the cell (E2) which is the at the top of Column E with 2, 5, & 4 in it. Now press the PERIOD Key and then the CURSOR-DOWN key twice. The second box should be high-lighted and you may press the ENTER key. The range specified should be: E2..E4

Immediately the computer will ask you to enter the OUTPUT RANGE. This is the location for the column of numbers which will contain the results of you multiplication. Press the CURSOR-RIGHT key twice (or wherever you want the results) and then press ENTER. Column G, on row 2 would be good. Immediately the results of the matrix multiplication of the first two boxes will appear in Column G on rows 2, 3, & 4.

Your results should now look like this:

3	2	1	MULTIPLY	2		20	Difference
4	3	2	TIMES >	5	EQUALS>	31	in products
5	4	3	THIS	4		42	11
Multiplicand				Multipliers		Products	

Notice that the difference in each of the numbers in the products' column is 11. The also happens to be the sum of 2+5+4.

Here is what is happening mathematically:

(3X2)+(2X5)+(1X4)=20		(A2*E2)+(B2*E3)+(C2*E4)=20
(4X2)+(3X5)+(2X4)=31	or	(A3*E2)+(B3*E3)+(C3*E4)=31
(5X2)+(4X5)+(3X4)=42		(A4*E2)+(B2*E3)+(C4*E4)=42

Each entry in each row is being multiplied by each number in Column E as the products are added together.

You should analyze how the numbers are being multiplied by altering the numbers in the first box and running through the matrix multiplication again and again. Read through the rest of this paper and then try the following combinations:

3	2	1	MULTIPLY	2		6	Diff. of
4	3	2	TIMES >	0	EQUALS>	8	2.
5	4	3	THIS	0		10	

And again:

3	2	1	MULTIPLY	2		12	Diff. of
4	3	2	TIMES >	2	EQUALS>	18	6.
5	4	3	THIS	2		24	

Let us examine the first example on the previous page. Essentially what is happening is that the number 3 in the first row of the first column (A2) in the first box is being multiplied times the number 2 in the first row of the second box (position E2). That result (6) is held in the computer's memory while the number 2 in the second column of the first box is multiplied times the number 5 in the second row of the second box. The second result (10) is held also. Finally (for the first row), the number (1) in the third column of the first box is multiplied times the number 4 in the 3rd row of Column E (the second box), and that result (2) is also held. Then the three results are added together 6+10+4=20, and the result (20) is placed in the first row of the box (cell position--probably Column G) you designated for your output range. For each row of the matrix (the 1st range, or multiplicand), you multiply by each entry in Column E (the 2nd range, or the multiplier).

Notice that if you change the numbers in the second box, the numbers in your output range will change also (just as they do if you change any of the numbers in the first box). Thus, if you wish to keep the results of any one of your experiments, you must re-specify the output range or location in order to leave the first results on the screen. It would also be practical, at the same time, to set up a second location (2nd multiplier) so that you can still see what it was (the 1st multiplier) you used to get your first results. In order to do this, you must press ESCape when the program asks you for the 2nd multiplier range, or

location. (Always press the PERIOD key at the new location before moving the cursor to define the range.)

ASSIGNMENT #4: CREATING A PERSONNEL DATABASE

PURPOSE: This assignment will give you an opportunity to use R:Base System V for the creation of a database that would be useful in a Ministry or other agency. The assignment will give you more understanding of relational databases in general as well as increasing your knowledge of R:Base commands. Those participants already familiar with R:Base are encouraged to develop more sophisticated applications suitable for the needs of their ministries.

DIRECTIONS: Working in pairs, create a database of education system personnel (teachers, supervisors, administrators, district or provincial national education officers, etc.). The general procedure you will follow is:

1. Specify the data elements needed in the database and design the overall structure (on paper).

2. Translate the desired structure into the appropriate computer coding for R:Base.

3. Develop (in input form) the computer-base data table.

4. Input example data.

5. Design format of output reports (on paper).

6. Develop at least one output report using R:Base.

More specific instructions will be provided in class and individual help will be available.

ASSIGNMENT #5: SYNTHESIS AND APPLICATION OF WEEKS 1 & 2

PURPOSE: This assignment will ask you to reflect on the content and skills you have gained over the last two weeks. Having come to the half-way point of the BRIDGES Workshop, it is appropriate to take a few minutes to reflect on how the material presented in class can be most beneficial when you return to your home country and position.

DIRECTIONS: As you think back over the concepts presented during the last two weeks (goals/objectives; educational modeling; population, enrollment, and financial forecasting; manpower planning; management information systems; etc.), consider the following questions:

1. Which of these concepts has the most relevance to your particular position and office in your home country?

2. What particular actions would you like to take to implement one or more of the concepts that have been presented?

3. What are the real-world barriers you might face in attempting this implementation? How can they be overcome?

4. What concepts would you like an opportunity to learn more about?

You do not need to write a response to these questions. However, please consider them carefully. They will be addressed during Friday's class time.

REFERENCES: "Educational Planning", Chapters 1-6.

Section III: Educational Policy Analysis

EDUCATIONAL POLICY ANALYSIS

by

William K. Cummings

Table of Contents

Chapter 1 **The Variety of Problems for Educational Research**

Chapter 2 Clarifying the Problem and Preparing the Research Design

Assignment #6: Identifying and Creating Variables.
Assignment #7: Univariate Analysis, Cleaning Data.

Chapter 3 **Considerations for Evaluating and Editing Research Reports**

Assignment #8: Using a Statistical Package.

Chapter 4 **Conducting and Reporting Multivariate Analysis**

Assignment #9: Conducting a Policy Study.

Chapter 1

THE VARIETY OF PROBLEMS FOR EDUCATIONAL RESEARCH

The more we study educational processes, the more their complexity becomes evident. It would be impossible and undesirable to focus on more than a small part of the complex educational world in a given research project. Still, before one particular research project can be discussed, it is useful to gain some picture of this complexity. The four questions below can help to clarify policy-oriented educational problems. Each question will be discussed in detail in this chapter.

WHAT IS THE SUBSTANCE OF THE PROBLEM?

Is it access, administrative inputs, classroom processes, or other?

WHAT IS THE SCHOOL LEVEL AND TYPE?

Is it a problem of access to the primary, lower secondary, or academic secondary level? Does the problem relate to higher education, some form of nonformal education, or a combination of these?

WHO IS THE RESEARCH FOR?

Is the study intended to help the national government, the local government, a particular institution, or some other client?

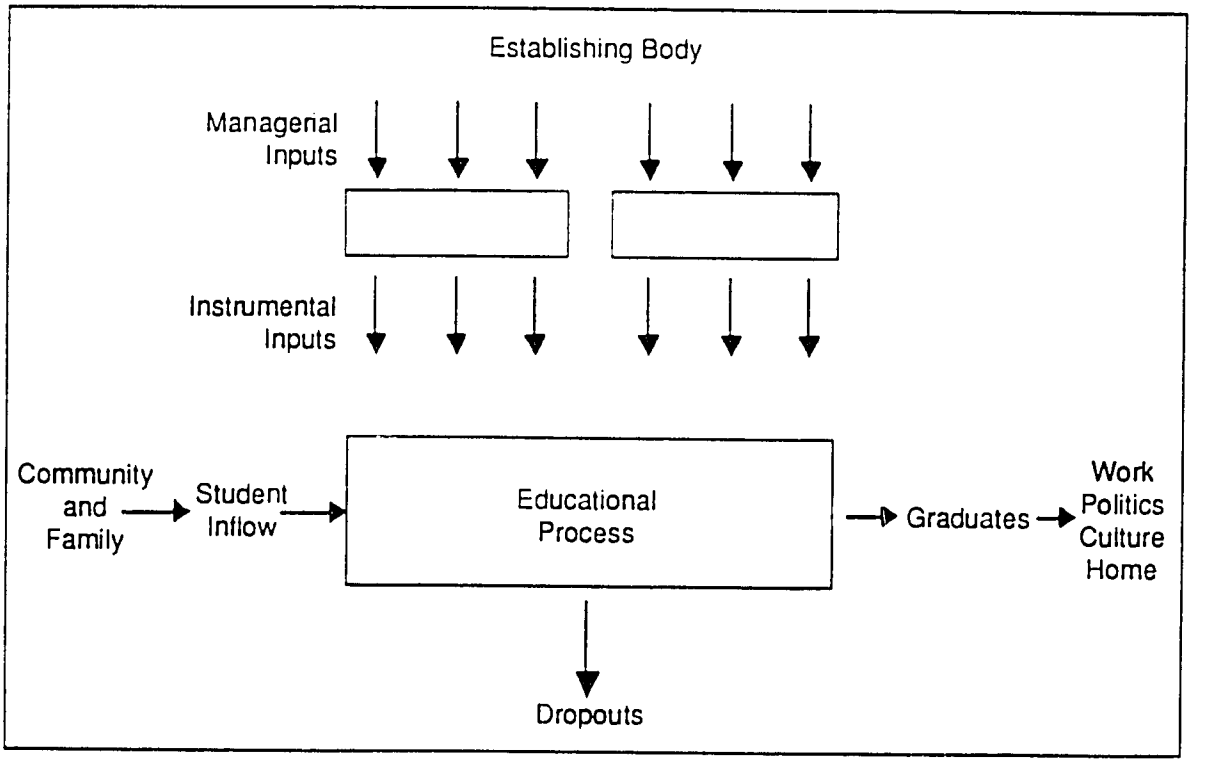
WHAT IS THE PURPOSE OF THE STUDY?

Some research is exploratory because it provides systematic information on a problem. Other research aims at guiding a particular policy decision. Exploratory research gives results that are not specific enough to guide a particular decision. Research designed to guide one particular decision will be too narrow to give much insight on the broader context. It is important to determine the purpose before the research design is finalized.

WHAT IS THE SUBSTANCE OF THE PROBLEM?

The simplest way to conceptualize the educational process is to draw from systems theory, as in Figure 1 below.

Figure 1.1: Schematic Diagram for the Education System

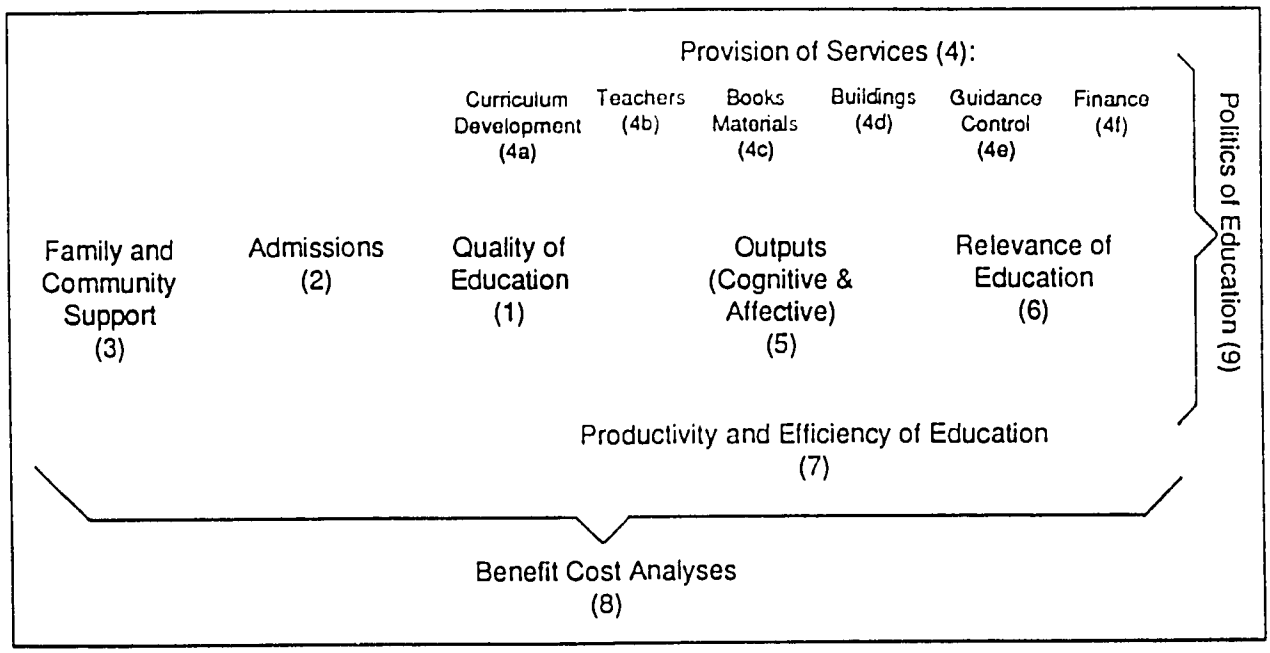


The educational process, or what goes on in schools, is viewed as taking place within a system. This process depends on inputs (mainly students) from homes and communities. Governments and other establishing bodies also provide inputs such as teachers, books, and materials to produce graduates. Drop-outs from the system are also part of the process. This simple figure highlights the items that educational planners are most concerned with, such as numbers of students, graduates, and teachers.

The figure also includes virtually any problem that an educational researcher or planner might wish to investigate. However, researchers and planners may be interested in different parts of the figure. For example, in looking at graduates, the educational planner is mainly interested in numbers. By contrast, a researcher might be more interested in the intellectual qualities and motivation of graduates. A planner is often uninterested in the black box of educational process, but this is a gold mine for many researchers. Similarly, while a planner is mainly concerned with the number of students entering the educational process, some researchers feel it is important to examine the family background of these students, the facilities and culture of the community from which they come, and so on.

The full range of topics suggested by Figure 1 will not be described here. However, it may be helpful to identify some broad categories of topics that are of interest to researchers and policy makers. The topics are listed in nine categories on Figure 2 below.

Figure 1.2: Grouping Educational Research Topics



1. Quality of Education

At the core of the diagram is the question of what actually takes place in the schools and classrooms. Researchers are showing increasing interest in questions such as these:

How much time do students actually spend on school work?

How often are students--as opposed to teachers--the center of action in the classroom?

Do the more able students help their peers?

Does the teacher relate equally to all students, or show favoritism to those that are more able, who dress better, or speak in the teacher's dialect?

These questions can be investigated in relation to input or outcome studies. But it is also worthwhile to investigate these topics in their own right. Very little is known about what goes on inside schools and classrooms. Still, there are many policies designed to transform this unknown. For example, extensive teacher training programs and modular instructional systems are all intended to create a "new" quality of education, yet these policies are based on a poor understanding of the "old".

2. Admissions

The question of admissions procedures is becoming increasingly important at all school levels. At the primary level, an interesting question is whether tuition charges have actually been abolished. At the lower and upper secondary levels, it is important to know whether schools use equitable admissions procedures, given the common excess of demand over supply. There are frequent rumors of bribes. What is the extent of this? How are the funds used?

In large cities, children are expected to attend the school nearest their home. Is this always true, or do children attend elsewhere? Is the ratio of applicants to places available at certain schools much higher than at others? Some of these questions can also be asked in in relation to the university level.

Researchers working at the university level are also interested in the high dropout frequency. Pre-admission methods to predict what kind of student will best be able to

3. Family and Community Support

The role of family and community support have received surprisingly little attention, possibly because they are somewhat difficult to affect through centralized policies. However, research shows the strong impact of family organization (e.g. number of children, mother's education, child-rearing practices, reading at home, parental assistance in homework) on school performance. Likewise, some communities develop a special educational climate which results in a disproportionate number of children advancing to higher levels of education. Family and community support are more easily mobilized by local leaders than by central policy actions. It would be useful to explore the importance of these resources and consider policies that might be appropriate for their mobilization.

4. Provision of Services

Research in this area is obviously important. Fortunately, there is already a modest volume of research on most services. One main exception is the lack of research on the true practices of financing schools. Schools are known to obtain extra income from parents to finance their operations. However, there is little information on the extent of this income, whether it causes an excessive burden on some parents, and whether this extra income is put to appropriate use. Likewise, are parent "contributions" related to the attention received by their child? These questions reflect on the reputation of education and deserve careful inquiry.

One important limitation of the current research on the provision of services is that it is usually authorized by the very bureaus providing the services. This is why the scope of inquiry is usually restricted. There is a need for some radical departures in research on the provision of services. The following questions should be asked:

Is it wise to reward all teachers throughout the nation according to the same pay scale?

Would it be best to provide incentives to teachers in remote areas?

Why is subject matter relevant to agricultural production so often absent from primary textbooks?

What are the pluses and minuses of various designs for school buildings?

How many children should study together in a school?

Until recently, surveys were the only means of answering these questions. However, in some countries, an expanded battery of questions is now included in school censuses which enable the provision of services to be evaluated. Secondary analysis of these statistics could prove very rewarding.

5. Cognitive and Affective Outcomes

Educators are naturally concerned with determining whether their plans and programs result in the desired cognitive and affective learning outcomes. They also need to know what variables affect the relative success of their efforts. Although measuring the cognitive outcomes of education is often difficult, some progress has been made.

An interesting finding from this research that deserves more extensive investigation is the relation of school size to student achievement. While current studies indicate that large size is related to higher average achievement, it is too early to conclude that small schools should be consolidated. Likewise, can the poor performance of children in small schools be due to the quality of education provided by the school or is it related to home factors? Answers to these questions are of special relevance to the outlying provinces that tend to have many small schools.

Another area of emerging interest concerns the success of schools in developing the moral and intellectual qualities appropriate for the new nation. Are young people concerned with development? Have they gained the values of the nation? What kinds of educational experiences are most effective in achieving the intended affective effects?

6. Relevance of Education

Most children attend school with the hope that it will enable them to cope better with the adult world. They also expect that it will help them to find a job. Underlying issues related to the curriculum and admissions procedures for each type of school are assumptions about their

occupational relevance. Are these assumptions valid? For example, do most graduates of the upper level Home Economics Schools become housewives, or assume jobs in occupations related to their training? Are Normal Schools and teacher training colleges the best sources for primary school teachers, or can academic track graduates also become effective teachers?

It is apparent that many young people spend a long time searching for jobs after graduation. Are some schools more effective than others in placing their students? Why?

Yet another area of investigation concerns the orientation of young people to their political responsibilities and to the task of national development. Are schools conveying the appropriate values, and do young people conform to public expectations in these areas?

7. Productivity and Efficiency of Education

One of the most firmly established traditions of research is the investigation of the productivity and efficiency of the educational system. Since the early Seventies, most Third World countries have annually collected data from every public school in the nation on the number of students, dropouts, graduates, repeaters, and promoters. This data has been analyzed at the national level to compute various indicators of productivity and efficiency.

In recent years, some attempts have been made to use similar indicators for appraising provincial level productivity and efficiency, and even the performance of individual schools. However, there is room for much more development of this research at the provincial level. Moreover, thus far there are few attempts to determine why provinces and schools vary in their productivity and efficiency.

8. Cost Benefit Analysis

Ideally, educational decision-making is based on knowing the likely results of spending limited resources in different ways. School officials naturally want the greatest benefits from their expenditures. Cost benefit research is especially useful when evaluating the wisdom of starting new programs. Unfortunately, in many countries, no research of this kind is conducted.

9. Politics of Education

Another unexplored area is the process by which decisions are made. Educational policy makers believe that their decisions are rational and assume that others in the educational system use the same criteria. But they also realize that they must take into account stubborn political realities when making decisions <u>if</u> they want to see implementation. What are these political realities?

For example, in a certain country, researchers might find that primary schools are built with impressive speed, but a new science curriculum is implemented only slowly. What accounts for the difference in the rate of accepting these innovations? Why are some schools or some teachers more receptive to a particular innovation than others?

To give another example, a study of teacher qualifications in one province of Indonesia showed that teachers in urban areas are far more likely to have the appropriate qualifications than teachers in rural areas. What are the patterns of decisions and incentives that produce these results?

In summary, there are many topics suitable for policy analysis. This section has listed nine broad areas of topics and many potential questions. Researchers should consider these areas carefully to be sure they understand how their projects fit into the overall scheme of educational policy. The next sections will discuss the remaining three questions: What is the school level and type? Who is the Research For? What is the Purpose of the Research?

WHAT IS THE LEVEL AND TYPE OF SCHOOL?

Once the substance of the problem has been identified, the next step is to identify the level and type of school. Often a research team settles for an important substantive topic (such as the efficiency of schools) without ever considering what level of schools they will investigate. In many instances, the team then decides to investigate all levels. A study across several levels could be interesting. Moreover, the field work involved in collecting data from multiple levels is not much greater than for one. However, there are drawbacks:

- The schools at the respective levels may be so different in character that comparisons across schools are impossible.

- Sampling frames and instruments designed for one level may not be useful at other levels. Hence, the preparatory work for a multi-level study is considerable.

- Limited resources provide a sufficient sample of schools at one level, but inadequate samples for each level in a multi-level study.

In most cases, the disadvantages outweigh the advantages. Thus it is usually best for a team to focus its investigations on a given school level. If there is a policy need for multi-level study, the team might extend its work to other levels after completing an initial investigation at one level. The first study will discover the problems that can normally be encountered in such an investigation. These problems can then be avoided in the studies that follow.

Some research problems involve the interaction between levels. For example, a research team might want to investigate why some upper secondary students choose to go to University, while some do not. Whatever the situation, it is always important to clarify and limit the focus of the project. One important way to do this is by clearly specifying the level and type of school.

WHO IS THE RESEARCH FOR?

Depending on how a given topic is approached, the results obtained may be of use to national leaders, to local leaders, or possibly to no one. The results of a study on the efficiency of the national school system would be of primary interest to central decision-makers. If the same study computed data at the provincial level, it might be useful to provincial decision-makers.

In a study of how science is taught in the schools, a researcher might focus on policy variables determined at the national level, such as the quality of texts and equipment. Yet this is a process largely controlled by provincial and local authorities. Researchers need to have a clear idea of who they intend their research to benefit as they plan their work. The nature of the beneficiary should have a strong influence on the variables that are selected for study. It is not enough for researchers to say they are doing a policy study; they need to have a clear grasp of whose policy they wish to influence. When the beneficiary is in focus, researchers need to spend time finding out precisely what policies the beneficiary is in a position to alter. These changeable policies should provide the key to the selection of key variables.

WHAT IS THE PURPOSE OF THE STUDY?

Policy research varies from very rough studies that explore a situation, to fine-grained experimental studies that attempt to clarify which, from among two or more policy alternatives, works best.

Researchers have to be clear about which kind of objectives they intend to pursue, for this decision will affect virtually every decision in the research design. Briefly, there are four types of design:

- Exploratory
- Correlational
- Causal-relational
- Experimental

These four kinds of research designs are described on the next page.

1. Exploratory Research

This is open-ended research that attempts to find the significance or meaning of a single concept such as Mysticism, the Mental Life of Delinquents, or the Role of a Secretary. At the beginning of the study, researchers have relatively few ideas about what they plan to look at, other than the focal concept. At the beginning of the project their measurement technique is likely to be unstructured observation.

2. Correlational Research

Researchers undertake this kind of study when they are interested in a single concept and wish to find out what is related to this concept. The research begins with a list of other concepts that could be related to this focal concept. Normally, the research proposes some structured techniques of measuring the concepts in the study. However, the researcher makes few assumptions concerning the relative importance of concepts or their causal-relational properties.

3. Causal-Relational

In this kind of study, the researchers want to determine whether a certain variable or variables cause or are caused by another variable or variables. This hypothesis is evaluated by measuring the strength of the relationship between the focal variables with data from a relevant population. In this evaluation, other variables thought to influence the relationship are controlled as much as possible.

An advanced form of causal-relational research is to evaluate a hypothesized model in which a complex of causal relations are specified.

4. Experimental Research

In this type of study, researchers propose to introduce a variety of treatments to a number of groups that are felt to be similar in all other respects. The goal is to determine whether the different treatments lead to different results.

The main features of each design are contrasted in the chart below:

	Central Focus on 2 or more Variables	Causal Order of Variables Specified	Values for Key Variables are Specified
Exploratory	No	No	No
Correlational	No	No	No
Causal-relational	Yes	Yes	No
Experimental	Yes	Yes	Yes

Chapter 2

CLARIFYING THE PROBLEM AND PREPARING THE RESEARCH DESIGN

The early steps in a research project are very important. It is not wise to make all the decisions about a research project by oneself or through discussion with a few colleagues. Rather, decisions should follow from a lengthy process of consultation and reading. Adequate preparation enhances the likelihood of producing useful knowledge.

The early steps in a project can be summarized under two main headings: (1) Clarifying the Problem and (2) Preparing the Research Design. Each area will be discussed in this chapter.

CLARIFYING THE PROBLEM

There are four steps researchers should follow in choosing and clarifying the problems for their research:

1. Consulting with Policy Makers
2. Reviewing the Literature and Available Data
3. Making an Exploratory Field Trip
4. Firming up the Research Strategy

1. Consulting with Policy Makers

If one is interested in conducting useful research, it is helpful from the beginning to establish contact with the policy makers one wants to influence. These individuals may or may not understand research, but increasingly they believe in research and want to take research into account when making decisions. For this reason, policy makers are surprisingly open to researchers. The research worker will be surprised at how easy it is to arrange an appointment with the appropriate government official.

Because the researcher's objective is to achieve influence, the best strategy is to try to meet with an official of high office. During the course of that appointment, the researcher should indicate a desire to meet others in the official's office and to have the privilege of reporting back from time to time as the project develops. The official will probably agree to both of these requests, and it will be wise for the researcher to follow up on them. The best predictor of the usefulness of a research project is the degree of understanding that relevant government officials have about the project. Ideally, the researcher is able to recruit one or more policy makers to his team.

The researcher should not expect to get a clear suggestion for a research topic from the policy maker. Policy makers usually are concerned with everyday decisions and seek information directly relevant to the here and now. The researcher will not be able to produce information so quickly. Rather, the conversation should focus on:

- Persistent problems faced by the policy maker,
- Problems the policy maker anticipates in the near future,
- New policies that the policy maker is considering or actually plans to implement, and
- Suggested educational problems or topics that could be researched.

In the discussion, the researcher should be open-minded and actively thinking about ways that research might help the policy maker. Usually, it is unwise to end a first meeting with a commitment to a particular topic. However, based on the initial discussion with a key policy maker, the researcher can follow up with lower-level people until two or more promising topics are identified. Then, another meeting can be set up with the key person for a confirmation of mutual priorities and interests.

2. Reviewing the Literature and Available Data

Once the researcher has some insight on the types of problems that are most relevant, the next step is to find out what is already known or knowable about these problems. On any problem of importance, a number of related studies may already have been completed. The researcher should make an effort to identify and obtain these related studies.

If possible, the researcher should also review the related international literature. Some good sources are the back issues of journals such as Comparative Education Review, Sociology of Education, Economic Development and Cultural Change, etc. Also, the researcher might get an interesting response by writing to one of the international documentation centers.

In reviewing completed studies, the researcher may discover some that have collected excellent data that could be used to answer the researcher's problem--even if the study report has a different focus. For example, a researcher interested in problems of school administration may find that a national study of school achievement has many questions on school administration. In such a case, the researcher might consider a secondary analysis of this data rather than spending much money and time to collect new data. Increasingly, researchers are coding their data so that it is easily accessible to others.

Yet other sources the researcher should review are school statistics and other routine statistics collected by local and central governments. These statistics are often highly relevant to policy studies. For example, they provide an excellent source for studying the incidence of drop-outs, the location of schools, frequency of student attendance, the training of teachers, and even some aspects of school facilities and finance.

Increasingly, central and local government offices are saving these statistics in formats that enable interested outsiders to carry out re-analyses. The advantage of these routine statistics is that they are collected from all the educational units of a given area, whereas most research projects depend on samples from all the units. Given the availability of these statistics, if they are sufficient to answer a policy question, it would be irresponsible for a researcher to collect the same data a second time.

In policy research, it usually makes sense to use school statistics to develop a general outline of the situation and to answer some of the main questions. However, for a full response to the problem, the findings from school statistics analyses need to be supplemented with additional data collected from the field. Choices of this kind can only be made after a thoughtful review of the available data.

3. Making an Exploratory Field Trip

Another valuable strategy is to visit a few of the types of people, schools, or communities that will ultimately be researched. In these visits, the researcher should try to get (1) local perspectives on the problem, (2) some sense of the information available at the local level, and (3) an understanding of local taboos.

Researchers working in cities often think they can seek any information they wish from local people, but exploratory visits may show otherwise. Likewise, local communities may follow different schedules than urban centers, or a visit may indicate the relative wisdom of relying on community leaders versus school principals for arranging field work. Generally, these exploratory visits can sensitize researchers to the issues and difficulties they may encounter as they conduct research.

4. Firming Up Research Strategy

While going through the steps above, the researcher will gradually gain an idea of what is possible and needed. At this point, it is time to settle on an appropriate research strategy. The substance, the school type, and the range of policy issues that should be addressed will all become clear. Finally, it will become evident who is most likely to benefit from the study. With this information, the researcher is ready to draft a research design. The section which begins on the next page will discuss the characteristics of a good research design.

PREPARING THE RESEARCH DESIGN

The drafting of a research design (RD) should not be approached lightly. Two important points should be considered in this regard:

- The careful drafting of an RD enables the researcher to logically integrate the various phases of the research process.
- The draft also enables the researcher to foresee the major difficulties that are likely to be encountered in the research process.

Careful preparation of an RD takes time--from a month to a year or more. It may result in a thick document. One RD prepared by an Indonesian team on the work experience of secondary school graduates was over 80 pages long; nearly 100 hypotheses were outlined and justified. Many problems can be handled in less space, although 3 to 4 pages will rarely be adequate. The types of questions listed below must be answered in a research design:

1. What is the title of my project?
2. What is my question?
3. Why is an answer to this question important?
4. To whom is this question important?

5. Who has tried to answer this question before? (Names, titles)
6. What were <u>their</u> answers?
7. What is my expected answer? (Hypothesis)
8. Why do I expect to find this answer?

9. How am I going to answer the question?
10. What do I need to observe to get the answer?
    - How many of X do I need to observe?
    - What characteristics need observation?
    - How will I measure these characteristics?
    - How will I collect data?
    - When is the best time to collect data?
    - When is it impossible? (Holy Days, holidays)
    - How will I organize data?
    - How will I analyze data?
    - How much time and money will be necessary?
    - Do I need permission? How will I get it?

A good research design should provide detailed answers to each of these questions, and each answer should be integrated with the others. The decisions to select the sample should logically come from the problem statement rather than from textbook notions about the desirability of stratified random samples. The personnel and budget should reflect the requirements of the project rather than the need to employ all the members of the chief researcher's academic department or government offices.

In principle, as each research problem is unique, so each research design should be unique. However, certain topics are generally included in most RDs. They are:

1. Title
2. Background of the problem
3. Nature of the Problem
4. Hypothesis
5. Definitions
6. Methods
7. Work Plan

Each of these seven topics will be discussed in detail. The major shortcomings and pitfalls to avoid in drawing up educational research designs will also be identified.

1. Title

While the title comes at the beginning of a research design, it is often wise to wait until all the other parts of the RD are completed before settling on the final choice of words. A good title should convey the true intention of the project interestingly, in as few words as possible.

2. Background of the Problem

The section immediately following the title is called the "background of the problem". In most RDs, this section is short and not especially informative. Yet this section is especially crucial and can achieve several objectives.

First, there might be an introduction to the project where, in one paragraph, the researcher outlines the key problem by stating

- o The importance of the issue,
- o The nature of previous work on the problem,
- o What questions are still unanswered,
- o What the project plans to research, and
- o What findings are anticipated.

The purpose of the introduction is to convince readers right from the beginning that the researchers know what they are doing, and hence, that further reading will be worth the reader's time. Many RDs fail to provide this introduction. Consequently, the reader immediately has to grapple with difficult technical and philosophical discussions. In these cases, readers are often discouraged and abandon further reading. Like the title, a good introduction is difficult to prepare until the research plans have been completed. If the introduction is written first, it may result in a section that is too long. Thus, the researcher is advised to save the writing of the introduction until the end.

Second, if the project has an applied objective, it is helpful to discuss the practical importance of the work. Statistics on the situation, official policy statements, quotations from informed people, and other current sources might be turned to as a way of illustrating why the research is needed. In some instances, this discussion can only report what is known. In other cases, it will be able to point to rationales for the research and potential causes that readers had never considered. Ideally, this discussion should stimulate the interest of practical people so that they will become active supporters of the project.

Third, the background discussion should review what is already known about the problem. This review should include a summary of operational knowledge about the problem, such as information the researcher gained from his interview with policy makers and preparatory field trips. An effective means of conveying this information is to present direct quotes from key people and from the researcher's field notes. Often the preliminary interviews will uncover significant differences in the operational knowledge of different people.

For example, a district official may state that all schools have books, whereas a principal may indicate that his school did not receive them until several months after the school year began. In an RD for a dropout study, one teacher might be quoted who thinks economic factors are crucial, while a woman teacher may say that parents do not equally support their daughters the way they support their sons. Each of these differences in operational knowledge suggests themes that should be included in the final report.

Finally, the background report or even a separate section entitled "Review of the Literature" should review and interpret the results of related research studies. For most research topics, there is already a substantial body of both indigenous and international literature on applied research problems. Yet many research designs never refer to even a single study conducted by another researcher. This neglect of established knowledge is inexcusable. The researcher should identify both what is valuable and what is probably mistaken or neglected in earlier studies. This interpretation then becomes the basis for pointing to new directions.

3. Nature of the Problem

In presenting the background of the problem, the researcher will have ample opportunity to suggest various aspects of the problem that need to be investigated. However, this discussion is likely to be fragmented and repetitive. Thus, a clear, compact statement of the nature of the problem can draw together all these facets. This statement might cover these headings:

1. The practical problem and that part which is researchable.

2. A list of the major related research efforts of the past and a summary of the relevant credible findings.

3. A summary of what past studies did not investigate, or investigated but failed to answer.

4. Finally, a statement of the aim, or what the proposed study intends to tackle. Remember not to try to do too much. The careful study of a limited problem area will be far more beneficial than an incomplete effort to examine a big problem.

4. Hypothesis

The previous section gives the background for the leading hypotheses of the study. But RDs often present the problem and even the hypotheses before they review the background and literature. How can someone know what should be done before examining what has been done? Research has to build on past experience. If not, the researcher is likely to repeat the mistakes of the past.

In the case of an exploratory study, the hypotheses will necessarily be few in number and simple in structure. For example, researchers studying teachers might expect to find that teachers in cities: (a) enjoy their work more than teachers in the country; (b) have better equipment to work with; or (c) have better living conditions and more friends. When these hypotheses are stated, researchers can then set some preliminary limits on what to study. Note the emphasis on "preliminary" because it may be impossible to identify all the big questions in a research design. A researcher may wish to revise these hypotheses once in the field.

5. Definitions

A hypothesis is a statement of an expected relation between two or more concepts. For example, a relation has been proposed between the location of teachers (urban or rural). At some point, the RD has to clarify what each concept in the hypotheses means. An exploratory study may only require a simple verbal definition to clarify. Correlational and experimental studies also need operational definitions and measurement procedures (discussed below). The verbal definition should be given in the hypothesis section, while the operational definitions can appear in the section on methodology. A helpful approach is to identify the key concepts of the study, provide their definitions, and then list the hypotheses.

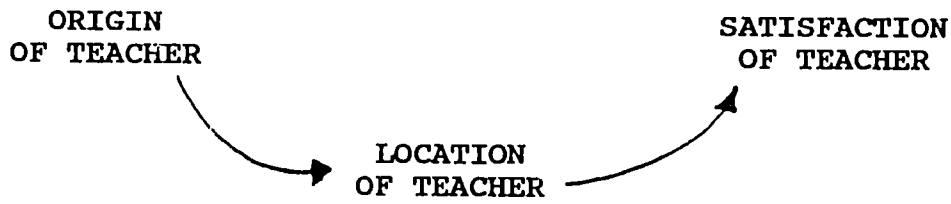
One frequent shortcoming of RDs is that they focus only on the bivariate relations between concepts, even when the situation may be more complex. While teachers who work in urban areas may be more satisfied generally, it is possible that teachers who were born and educated in rural areas may actually get more satisfaction from teaching rural children. These possibilities could be formulated:

For teachers of rural origin, teaching in rural areas is related to greater satisfaction.
For teachers of urban origin, teaching in urban areas is related to greater satisfaction.

At this level of complexity, the concepts can be divided into three groups that correspond to three kinds of variables. One group includes the dependent variables, which the study is trying to account for. A second group is the independent variables, which have obvious and direct relations with the dependent variables. Finally, a third group refers to those variables that introduce special conditions (such as origin of teachers), or may even explain the expected relations between the independent and dependent variables. This group can be called conditional or background variables.

Along with these lists, a diagram is also helpful, showing the conditional variables to the left, the independent in the middle, and the dependent on the right. Arrows drawn between the groups can be used to represent the researcher's assumptions about the direction of influence. An example of a diagram is on the next page.

Diagram of Causal Relationships



With experience, researchers learn that there are certain technical conventions to follow when preparing these diagrams. The use of diagrams will vastly improve the researcher's effectiveness in conveying the research plan to the lay reader of his RD. As the saying goes,"a picture is worth a thousand words".

In an exploratory study, it may not be necessary to discuss multivariate relations. Nevertheless, a proper correlational or experimental study must consider the complexity of relations between the concepts to be included in some depth. For example, the social background of a

student is known to have some effect on how a student behaves in school. Thus, to fail to take student background into account when considering aspects of student behavior is to fail to execute a complete study. The question of which complex relations to consider is best answered by the educational theorist. The issue of complex relations is crucial simply because when they are omitted from a research design, they may indicate serious shortcomings in the way the researcher thinks about the problem.

6. Methods

Along with the problem statement, the section on methods can be the most informative section of the RD. Here, researchers have the opportunity to demonstrate that they have carefully reflected on the practical aspects of the research process, and that they have mastered basic research techniques. Detailed work that outlines methods enables other professionals to provide critical advice that often will ease and improve the research process. Especially for a correlational or experimental study, the proper discussion of methods can easily exceed 20 pages.

Unfortunately, few researchers devote appropriate attention to the discussion of methods. RDs often have only one to two pages on methods, and the discussion is usually so unclear that it is meaningless. For example, when discussing the sample, an RD might say, "Ten children each from the third class of 20 middle schools chosen by the heads of each community in the district". This statement indicates that a number of critical sampling decisions have been made, yet it provides no rationale for these. In many RDs, it is impossible to find even a hint of the sampling rationale. Elsewhere, under techniques of statistical analysis, the RD may simply say, "chi-squared" or "correlations" without any further elaboration or justification.

For the contemporary researcher, the potential benefits of carefully defining the methods outweigh the losses. Hence, researchers should be as clear and precise as possible. The discussion which follows presents five weaknesses commonly found in the methods sections of RDs. They are:

A. Failure to specify the unit of analysis
B. Lack of focus in data collection
C. Inadequate selection of sample for problem
D. Inadequate operational definitions
E. Failure to identify appropriate techniques of statistical analysis

statistical analysis

A. Failure to specify the unit of analysis:

One of the most common weaknesses is the failure to clarify the unit of analysis. Educational researchers are usually interested in something that goes on in schools, but they rarely clarify in their RDs whether they intend to study the student, the teacher, the principal, or the school as a whole. Whatever the problem, any one of these actors could be the unit of analysis.

Nevertheless, it is difficult to focus on all of these potential units in a single study. A researcher's decisions about the sample chosen, the ways in which the variables are to be measured, and tables are to be constructed, all relate back to the unit of analysis. A focus on students will lead to an entirely different set of decisions from a focus on teachers. Thus, a researcher's first job is to clarify which actor is the focus of his study.

B. Lack of focus in data collection:

The school as a research site is tempting. While the teacher may be selected as the unit of analysis, it will be possible once at the site to collect data from many other actors at the same time--e.g. the students, the principal, local community leaders, an observation sheet concerning equipment at the school, and so on. The researcher's first reaction in the face of this abundance is to ask, "Why not?" A more thoughtful researcher would ask, "Why?"

The costs of collecting data from multiple sources are:

- Field work takes more time
- Given a fixed budget, the sample size for the basic unit has to be decreased if multiple sources are planned
- The school is more disrupted when multiple sources are planned
- The researcher's data analysis is more complicated.

The researcher should take all of these factors into account before deciding to collect data from more than one source. Once these costs are reviewed, the researcher may decide that a single source provides enough information.

If a research team decide to collect data from more than one source, then they should always keep in mind that each source will be approached only to produce insights on the focal unit of analysis. Thus, if the unit of analysis is the teacher, the sampling design will be developed to create generalizations about teachers. The researcher may also collect data about students, but should not attempt to make generalizations about students. Rather, the student data should be aggregated to produce generalizations about the student attributes of teachers in group A, group B, and so on. Many attributes of students (such as age, family background, whether they write with their left hand or not) may be of interest as separate topics. But since these attributes will not tell anything about the teacher, who is the focal unit of analysis in the project, it is a waste of time to measure them. The researcher must always keep focused.

C. Inadequate selection of sample for the problem:

Most researchers are unable to examine all the cases relevant to their problem, so they select samples. Ideally, they should select a sample that fits their problem. For example, if the goal is to estimate mean scores for some attribute of a group, it is important to ensure that the cases are randomly selected from the universe of that group. First, the universe should be clearly identified. Then a proper sampling procedure should be identified and justified.

If two or more groups are to be repeatedly compared, it is important to select enough cases from each group so that the generalizations for each group do not have an excessively large sampling error. First, researchers should identify their target. Then they should indicate how they will make sure that enough cases are obtained from each group. When one or more target groups are underrepresented in the universe, a proportional design may be required.

If a multivariate analysis is to be used, it is important to have enough overall cases so this is possible. The greater the number of variables that are to be analyzed simultaneously, the larger the number of cases needed. Different research problems require distinct sample designs. In an RD, the logic relating the problem and the sample should be spelled out. Unfortunately, many RDs choose a sampling design that is unrelated to their purpose. The most common error is to rely on a cluster design regardless of the problem and the relevant universe. Such a cluster design approaches selection as follows:

1. A number of provinces are selected from all provinces.
2. A number of districts are then selected from each province.
3. Within each district, communities are selected.
4. Within each community, schools are selected.
5. Within each school, certain teachers or students are selected.

This cluster design is the worst design for virtually all problems an educational researcher might investigate.

D. Inadequate operational definitions:

The section on hypotheses should provide a more or less complete list of the concepts to be examined. The methods section should explain how each of these concepts will be operationalized. A useful format for this purpose is to prepare what is often called an operational matrix with the following headings:

OPERATIONAL MATRIX

No.	Major Concepts	Sub-Concepts	Source of Data	Operational Definition	Hypotheses Where Used
1.					
2.					
3.					

This matrix can be completed in whatever detail the researcher considers necessary. The elaboration of Operational Definitions is one area where we feel researchers should exercise special care.

Several advantages to developing a matrix are:

- To ease the researcher's task of preparing measurement instruments.

- To clarify instances where similar data are to be drawn from multiple sources.

- To clarify the researcher's operational plan, facilitate friendly critics to offer suggestions. For example, the researcher may propose to develop an instrument for measuring student attitudes to science. A friend at another institution who reads the plan, may know of a standard procedure for measuring these attitudes and may send a copy to the researcher.

Developing an operational matrix takes some time, but once developed, it speeds subsequent stages of the project. Most researchers who have developed operational matrices believe that it improved the quality of their project.

Yet another area that is usually overlooked in RDs is the planning of measurement procedures to ensure validity and reliability. Ideally, the section on methods should cover such issues as the procedures to be used in training field workers, the gains expected from the trials, and the other steps planned to improve reliability and validity.

E. Failure to identify appropriate techniques of statistical analysis:

When preparing an RD, it is not always easy to identify all the analytical techniques that will be used. On the other hand, a researcher who writes only "chi-squared" or "Pearson product-moment correlation" has apparently not put much thought into the analysis.

As with sampling and other aspects of methodology, different research problems will require different methods. Three common weaknesses are:

1. The excessive freedom with which data is aggregated into complex scales.

2. No consideration of the distributions of individual variables, a procedure called univariate analysis.

3. The lack of multivariate analysis.

3. The lack of multivariate analysis.

Most analysis is restricted to the inspection of relations between sets of two variables. Far too much is made of those relations that are discovered as well as those that are rejected.

Based on a sound grasp of the data to be collected, researchers should provide a few hints on their analysis strategies. For example, if they plan to use cross tabulations mainly, they might present a sample table and identify the types of inferences to be drawn from the table. The main point of this subsection is to present some evidence that serious thought was given to analysis strategy.

Many researchers conclude that the appropriate way to demonstrate serious thought is to write out the lengthy formulas used to compute statistical indicators, such as the formula for gamma, or the Pearson-product-moment correlation. Producing such formulas is not necessary; the reader can look them up in a standard statistics textbook. Formulas should only be written out when they are unknown to other researchers and are especially critical to the success of the project.

Work Plan:

In this section, the researcher details the team, budget, and schedule. Several suggestions are outlined below.

1. When the team is identified, individuals who will not make a genuine contribution to the project should not be listed. Above all, such individuals should not be rewarded with honoraria.

2. The team should be selected to get members who can effectively carry out the project, and not on the basis solely of personal friendships. One problem that most RDs seem to ignore is the qualifications of field workers. For some research projects, students or university staff will not be the best people. Housewives or teachers might be preferable.

3. Individuals from the institutions that will use the research should be included as active members of the team.

4. The budget for the RD should realistically reflect the costs of the project. The large funding requested by researchers for the small amount of data they intend to collect is sometimes amazing. Items such as honoraria, per diem for teams to visit sites, and budgets for consumption tend to inflate budgets. Ideally, the researcher's first goal should be to collect the data.

The guiding principle in determining costs should be to ask for the minimum amount of money necessary to collect the maximum amount of data and complete a satisfactory analysis and report.

5. The object of applied research is to inform policy. Hence, it is reasonable and desirable for the budget to include funds for a workshop where the results are shared with users and other research colleagues.

6. Sufficient funds should also be requested to cover the cost of producing multiple copies of the final report. The schedule should be prepared with careful attention to the rhythm of the research site. Intended dates for field work should take into account school holidays, tests, and other such unusual events.

Conclusion:

The researcher should attempt to draft an RD that fits the problem. Especially if researchers intend to explore a problem area, their design will take on a very different shape than either a correlational or experimental study. In the exploratory study, the problem discussion will be based mainly on interviews and observation, and the section on methods will be briefer and more qualitative. Still, it is reasonable to ask a researcher undertaking exploratory work to respond to each of the sections. An exploratory study that begins with a list of hypotheses, however vague, is far more likely to clarify the situation than one that lacks the initial focus. Thus, the comments of this chapter are relevant to all research problems.

A researcher is likely to prepare the RD in stages, perhaps first writing the problem statement, then moving on to hypotheses and methods, and finally returning to the title and introduction. After completing the various sections and assembling them, the researcher should reread the whole and ask the following questions:

- Is the RD well-written?
- Does it answer all the questions listed on page 5?
- Are there sensible justifications for each of the decisions in the RD: the sample, the sources of data, etc.?
- Do the different sections of the RD fit together?
- Finally, is the research really going to be of use to anybody?

Each of these questions needs to be answered honestly. In most cases, the first draft of the RD is discarded. Interestingly, one of the reasons why first drafts are so poor is that they are produced by a team of researchers. Each member is responsible for a separate section. Due to inadequate discussion among team members as well as differences in their skills and interests, they rarely create a well-integrated first draft together. Ideally, as second and third drafts are prepared, the above questions are answered clearly.

Chapter 3

CONSIDERATIONS FOR EVALUATING AND EDITING RESEARCH REPORTS

This chapter will present guidelines for evaluating and editing research reports. The chapter is divided into four main sections:

1. General criteria for research reports.
2. Common weaknesses in research reports.
3. Appropriate variations in emphasis of reports.
4. Improving the quality of reports.

GENERAL CRITERIA FOR RESEARCH REPORTS

The following are some general criteria for research reports which will be elaborated in greater detail as we turn to the respective sections of a typical report.

1. The research should attack a clearly defined problem, with a narrow focus. Throughout the report, the researcher should concentrate on the core problem.

2. The issue should be important for policy or science.

3. The report should be written clearly with few technical terms and numbers. If the report is statistical, representative details and interesting results can be included. Appendices should be used to present this information.

4. The report should provide enough information so that a new research team can either check the findings or replace the original team's research plan from beginning to end.

5. The report should be distributed to many libraries or indicate a convenient way for readers to get copies.

COMMON WEAKNESSES IN RESEARCH REPORTS

The content of research reports will vary according to the type of research problem investigated. Figure 3.1 lists the basic sections of all reports and point out typical flaws to avoid.

Figure 3.1: Common Weaknesses in Research Reports

Section	Typical Weaknesses
1. Title	Usually too long, unattractive.
2. Abstract	Usually not prepared; if provided often *too* long yet lacks balance because over-emphasizes either method or results. 250 words is enough.
3. Table of contents	Usually gives only routine headings, failing to suggest the special content of the report.
4. Table of tables and figures	Usually absent.
5. Background of the problem	Often neglects to suggest the social or academic significance of the problem. Literature review fails to cover related research; where literature covered discussion often lacks a clear relation to the report's particular research problem.
6. Clear statement of the problem to be researched with definition of key concepts, their expected relations, and the theoretical implications of these expectations.	Often definitions of concepts are not useful; usually report fails to indicate the reasons for expecting hypothesized relations; theoretical discussions invariably absent.
7. Brief history of the study providing opportunity to be opened about changes in plans necessitated by research conditions.	This section never found in reports; is it possible that all research proceeds smoothly as planned?
8. Summary of design	
a. Location of field work	• Often *no* explanation given for why one place chosen over another: actual reason hopefully because site is optimal for problem. Another good reason is transportation.
b. Sample	• Sample design invariably *lacks a rationale*, and is often wrong. Researcher must clarify unit of analysis and the critical variation in this unit to which sample is designed to include.
c. Instruments and tryout	• Often discussion fails to show how the indicators in the instrument are actually indicators of key concepts. Tryout rarely executed, and even when done the resultant changes are not reported.
d. Account of field work	• Sometimes timing is crucial; does study happen at optimal time?

9. Analysis and discussion	
a. Plan of analysis to help reader through the numbers to come. Can include some commentary on new statistics with examples.	• Usually lacking. When included, usually only verbal with no graphical illustrations.
b. Univariate analysis	• This is often absent, but successful bivariate analysis depends on understanding of component variables. Univariate analysis can lead to conventions for handling missing data, for collapsing multi-category variables into fewer number of categories. Also can lead to normative interpretations.
c. Scaling procedures	• Often unclear how indicators combined. Unfortunately, too often the procedures involve complex but unjustified weighting schemes. Also indicators often summed without prior examination of their interrelationship.
d. Bivariate analysis	• Design for tables often inconsistent. Graphics rarely included. Too many tables included in text that, since they have little relation to argument, should be included iun appendix.
e. Multivariate analysis	• Critical bivariate relations should be examined to see if they hold up when other variables are considered. Surprisingly this rarely executed.
f. Discussion	• Ideally, the analysis should be confined with a discussion that tells about each result and how it relates to the focal research problem. The analysis and the discussion should run together rather than in separate sections. Analysis should not report all results, but only those that make up the story.
10. Interpretations and conclusions	Too often we find unjustified interpretations. The study's limitations are forgotten, and conclusions are reported which do not follow from the study.
11. Bibliography	Often references incomplete so reader cannot trace.
12. Appendices a. Sample details b. Copies of instruments c. Listing of basic data d. Details of scaling e. Additional analysis or field notes f. Personnel and other administrative material	Little of (a) to (e) ever included.

APPROPRIATE VARIATIONS IN EMPHASIS OF REPORTS

Most research reports include practically all of the sections listed in Figure 3.1. It is preferable to allow researchers flexibility to follow the format appropriate to their particular problem. However, some variations in emphasis can be expected, according to the type of study conducted.

On the basis of the scheme presented in Chapter 1 to classify research topics, expected variations in emphasis can occur in exploratory, experimental, correlational and causal-relational research:

1. Likely Variations in Exploratory Studies

The exploratory strategy is used when relatively little is known about the problem's background or about relevant literature, or when the researcher is uncertain about whom to interview or the range of variables to consider. A report of exploratory research would have relatively little to say in the several initial sections outlined in Figure 3.1. However, as the researcher turns to the account of the field work (Figure 3.1,8d), analysis (Figure 3.1,9), and interpretations and conclusions (Figure 3.1,10), the report will become much fuller.

According to the problem and strategy emphasized in research, there will be some natural variation in research faults. In the exploratory study, the greatest danger is overgeneralization of results, because the sample base is small and likely to be biassed. Style and clarity of writing are particularly important in exploratory studies.

2. Likely Variations in Experimental Studies

The report of an experimental study gives a well-established background of past work and usually some carefully articulated theories. The early sections of a report on experimental research are likely to be long, while the sections on analysis, interpretation, and conclusions tend to be brief. This is because the early sections will narrow the focus, so the actual research will only include data on a few variables from two or a handful of small experimental groups. Experimental studies often make mistakes in the beginning because they misread existing literature, and thus create theoretical differences where

3. Likely Variations in Correlational and Causal-Relational Strategies

In these studies, report weaknesses are most likely to occur in the middle, with the selection of samples, preparation of instruments, development of scales, and analyses. Researchers tend to make unjustified decisions. They fail to think through the logic of their samples. A dropout study surveys only dropouts and thus is unable to compare dropouts with a group who stay in. Scales become enormously complex, combining 30 or more indicators that measure a common concept. Yet a careful reanalysis would show that many of these indicators are unrelated to each other. Analysis takes up all the possible combinations of variables, while forgetting the purpose of the analysis.

A much more systematic analysis of the typical weaknesses of different strategies could be given. But such an analysis would not be very useful. When a report is read carefully, the weaknesses usually become apparent. Once the weak points have been identified, the question to ask is: Are they minor weaknesses or fatal weaknesses?

IMPROVING THE QUALITY OF REPORTS

Because there is not much that can be done about bad reports once they are submitted, "preventive medicine" should be considered. What can be done to increase the chances that reports will be good? The following section provides five suggestions for improving report quality.

1. Establish a set of general criteria for what is expected out of a research paper. The beginning of this chapter lists the major, general criteria for quality reports. To this list other criteria might be added:

The importance of the problem to scholarship or policy makers.

The number of researchers involved in the project and their preparedness.

Whatever the criteria, the important thing is that the criteria exist, and that researchers know they exist. If researchers know there are criteria, they will be able to follow them.

2. Insist that researchers prepare a detailed research proposal, or design, prior to launching the project. Some proposals are so short and imprecise that they do not enable the researcher to do anything. For exploratory research, it is difficult to prepare a detailed design, but for all other strategies, it is feasible. Moreover, most research follows one of the more complicated strategies.

3. Establish a review board to critically examine all research proposals prior to starting projects. This board should have the authority to reject weak proposals, or request that these proposals be improved. Ideally, the review board should be complemented by some organization that can provide technical expertise to those research teams that prepare proposals. No research project should be allowed to proceed until a sound proposal (design) has been prepared. The surest guarantee of a quality research report is a quality proposal.

4. Establish some means to provide researchers with the key literature relevant to their projects. Researchers usually have access to small libraries that lack important references. Yet good research depends on a good understanding of the background of the research problem. Either the Research Center should send literature on each project to regionally-based teams, or a representative from those teams should be invited to the center to conduct a review of the literature.

5. Finally, provide adequate research budgets for teams to open workshops where qualified critics can review their work at various stages. Dialogue with other researchers during a project is the best way to discover weaknesses and correct them before the final report is written.

Chapter 4

CONDUCTING AND REPORTING MULTIVARIATE ANALYSIS

INTRODUCTION

In applied educational research, some variety of multivariate analysis is required if the researcher truly wants to demonstrate the importance of policy variables. Whenever a researcher wishes to show that a certain policy variable influences some desired outcome, other variables that also influence this outcome must be taken into account. The questions raised are:

Is the policy variable's influence strong or weak?

Does the policy variable have an influence independent of other relevant factors, or

Is the policy variable's influence really an artifact of the influence of these other factors?

For example, increasing the educational level of teachers should improve their performance. To support this view, a study could be conducted to show a relation between teacher's educational level and performance. But better educated teachers tend to work in higher quality schools that in turn, receive relatively more able children. Thus, to truly demonstrate the importance of teacher education, these other relevant variables have to be taken into account or controlled. In sum, a multivariate analysis needs to be conducted, when several variables are considered at the same time.

In policy analysis, there are a few situations where the researcher is not concerned with relations between variables, such as a study that simply tries to see if books have reached the schools. Hence, multivariate analysis is not required. However, in most cases, the focus is on relations between variables. Given the complexity of the educational system, this implies <u>multivariate analysis</u>.

While multivariate analysis is normally required, remarkably few studies ever attempt it. Some of the reasons for this are as follows:

1. Most researchers either do not know about multivariate analysis, or have never had the opportunity to systematically study the available techniques.

2. Those who have studied the techniques theoretically are reluctant to try them because they require extensive calculation.

3. Similarly, researchers are often overwhelmed by the seeming variety of multivariate techniques and uncertain as to which ones will be most appropriate for their purposes.

4. Finally, researchers are uncertain about the appropriate format for reporting the results of their multivariate analysis so that other researchers will approve of their work and policy makers will understand it.

This chapter will respond to each of these obstacles. While the chapter will not teach multivariate analysis, it will identify several of the more clear presentations of these techniques. Obviously, multivariate analysis requires extensive computation; if done by hand, this can be very time-consuming. However, at least for certain techniques, multivariate analysis builds on the computations required for bivariate analysis. Thus, it is not that much more burdensome. In any case, computers are rapidly becoming available to researchers and should reduce the significance of the time-burden. The focus will not be on the full-range of multivariate techniques, but only on those that seem practical for the near future, given current trends in computer development. The main thrust of the chapter will be:

1. To suggest an approach for preparing data that maximizes flexibility as the researcher selects an appropriate technique.

2. To point out the basic similarity of all multivariate techniques.

3. To provide guidelines for the selection of the technique most suited to each problem.

4. To provide guidelines to report the results of analysis.

GOALS SUITED TO MULTIVARIATE ANALYSIS

Before launching an analysis, it is expecially helpful to specify a preliminary goal. This is because the research goal as well as the measurement level of the variables in the analysis will influence the choice of multivariate techniques. The following are the most common goals pursued in multivariate analysis.

1. Confirming a Relationship

The researcher has found an empirical association between two variables, but is worried that this may be a spurious artifact of other relations. In applied research, one of these variables would be a policy input and the second, a policy output. Thus, it is necessary to control for other variables in an effort to increase certainty about the stability of the key relationship.

2. Analyzing Groups

The researcher expects that 2 groups (usually one of which has been exposed to some policy) will achieve different levels of performance on a common task, and wishes to make sure that the difference is due to differences in the nature of the two groups, rather than due to some external influence.

3. Identifying the Most Important Variables

Often a large number of variables are known to have some relationship to a given dependent variable, yet the researcher wishes to know which have the strongest relation. Especially, which variable is strongest, even when others are jointly allowed to influence the situation? Policy variables, hopefully, are among the most important.

4. Analyzing Context

If certain variables that are part of the normal context are thought to "contaminate" the dependent variable, the influence of these variables can be removed before determining which are most important. Alternatively, they may be taken into account in some other way.

5. Distinguishing Groups

A special case of Situation 3 is when the researcher is interested in finding out which, from among a long list of variables, effectively discriminate the members of one group from another.

6. Defining Causal Explanations

Sometimes the researcher is not only interested in finding out which variables have the closets relation to a dependent variable, but also in finding out the system of links between these variables--as well as between them and the dependent variable. This is a case of trying to construct a causal model of all relations.

7. Predicting Outcomes

Yet another goal of research is to combine several variables in a common equation to predict some outcome.

8. Drawing Inferences

The significance of the results is related to all of the above goals. In some cases, significance is not a special issue because the data are from the population. Inferences are sometimes less important than description. However, in most research situations, the aim is not only to describe a situation, but also to ensure that the description is significant. Inferential statistics appropriate to multivariate analysis serve this purpose.

A number of different multivariate techniques are useful for reaching the above goals. Table 4.1 on the next page provides a tentative summary, taking into account the implications of the variable measurements included in an analysis.

Table 4.1: The Utility of Various Multivariate Techniques

	Measurement Level					
	Dependent: Qualitative			**Dependent:** Quantitative		
	Independent are:			**Independent are:**		
	Qual	Mixed	Quant	Qual	Mixed	Quant
GOAL						
1. *Confirm a Relationship*	Partial Gamma	←	←	→	→	Partial Correlation
2. *Analyze Groups*	Partial Gamma	Discriminant Analysis	←	ANOVA MCA	Multiple Regression	→
3. *Identify Most Important Variables*	Log-linear Models	←	←	ANOVA MCA	Multiple Regression	
4. *Analyze Context*	Log-linear Models	←	←	→	ANCOVA Multiple Regression	
5. *Distinguish Groups*	Discriminant Analysis					
6. *Define Causal Explanations*	?	?	?	Commonality Analysis Path Analysis		
7. *Predict Outcomes*	?	?	?	Multiple Regression		

NOTE: → = some data need to be transformed before analysis is possible

= the cell is not relevant

? = techniques are available but not very effective

From this table it is clear that:

1. The techniques for the first goal are different than those for the remaining goals.

2. For goals 2-7, multiple regression is the most frequently mentioned technique. For this reason, the researcher is advised to seriously consider using this technique in planning analysis.

While it is not apparent from the table, these goals and techniques can be thought of as being ordered in a hierarchy. It is reasonable for the researcher to begin at a relatively low level in this hierarchy, and gradually move up. The original research goal may simply be to confirm a relation, but ultimately, researchers may construct a causal model. As they move up, they are likely to make changes both in the measurement level of the constituent variables, and in the techniques employed. To anticipate this progression, researchers should prepare the data for maximum flexibility.

PREPARING THE DATA

Before starting the journey into multivariate analysis, it is most important that the data be put in order. Most studies collect data at both the qualitative and quantitative levels. The steps below can reduce the difficulty of working with this mixed data.

Steps to Preparing Data

1. Prior even to collecting the data, the researcher should prepare a variable matrix that specifies the central concepts, their presumed temporal and causal ordering from more independent to more dependent, and the procedures to be used to measure these variables.

2. Once the data are collected, they need to be cleaned, and useless data are dropped.

3. All qualitative data should be carefully inspected. If possible, new dichotomous variables should be created from each of the original variables. In other words, from a single original five-category variable, such as religion, can be created the new dichotomous (dummy) variables:

(a) Islamic: Yes or No.
(b) Christian: Yes or No.
(c) Hindu: Yes or No, etc.

This procedure of collapsing, to create new, dichotomous variables, should be repeated for all qualitative variables.

4. Medians, means, and standard deviations for all quantitative variables should be computed.

5. Where the distribution of a particular quantitative variable seems too skewed, a new more normally distributed variable should be created by using an appropriate transformation.

6. Create a new qualitative variable out of all the quantitative variables by assigning a score of 0 to values below the median, and of 1 to values above the median.

7. If several indicators were used to measure a common concept, they should be cautiously examined. Some or all of them, if justified, can be combined into an index or scale.

8. Create a matrix that indicates the bivariate relations of all the variables in the study. This can be done either of two ways:

- By computing simple gammas between all of the dichotomized qualitative and quantitative variables, or more preferably,

- By computing Pearson product moment correlations between all of the dichotomized qualitative variable and the original quantitative variables.

With experience, researchers will see the value of carrying out all or most of the steps mentioned above before moving forward. In multivariate analysis, most of the information created in these 8 steps will be necessary at some phase. The next section of the chapter will describe the three basic groups of multivariate techniques.

AN OVERVIEW OF MULTIVARIATE TECHNIQUES

Once beyond bivariate analysis, the large and growing diversity of sophisticated techniques available can be confusing. Each technique has a special name and an apparently unique formula. Each has a special purpose, and each produces its special set of statistics. It is quite natural to feel overwhelmed by this diversity.

In this chapter, all the techniques mentioned in Table 4.1 will be discussed 4.1 except for log-linear models (which may not yet be feasible for widespread use).

The techniques listed in Table 4.1 can be divided into three groups:

1. Relational techniques of partial gamma, partial correlation, and analysis of variance.

2. Basic techniques of multiple regression, discriminant analysis, N-way analysis of variance, and analysis of covariance.

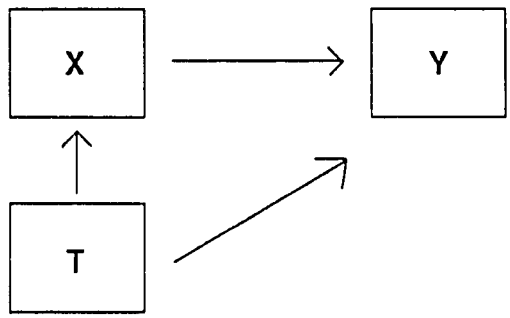
3. Derived techniques of Multiple Classification Analysis, Commonality Analysis, and Path Analysis.

The major divide in these three groups is between the first and last two. The relational techniques are only concerned with confirming the importance of a single independent variable. They provide no special information on the role of other dependent variables, even those involved in the confirmation process. By contrast, the basic and derived multivariate techniques produce information on all the variables involved in the analysis. Because the two relational techniques are relatively easy to compute and understand, they are considered first.

1. Relational Multivariate Techniques

Three-variable analysis is the simplest level of multivariate analysis. In certain situations, it is sufficient. Unlike more complex levels, the computations can be reasonably carried out by hand. Three-variable analysis illustrates several of the common problems of multivariate analysis, regardless of level.

The situation most appropriate to three-variable analysis is when the researcher has found a meaningful and significant relation between two variables, such as teacher's education (X) and student performance (Y), but knows that other variables such as school quality (Q) and urban location (L) are related to both focal variables. In such a case, the researcher might draw a simple diagram as follows:



Let T=Q, L, etc.

A simple way to test for the importance of the X-Y link is to perform a series of three-way tests where different substitutes for T become the control variables. Provided that the relation between X and Y remains statistically significant through each of these tests, the X-Y relation can be assumed to be important.

Three rules should be followed in selecting variables for the control situation (i.e. T).

1. Only those variables that logically and temporally precede Y should be considered.

2. Only the variables that can be thought to have some impact on the X-Y relation should be considered.

3. Only the variables that have a substantial association with Y should be considered.

These three rules can limit the number of variables that are actually subjected to the three-variable test.

As soon as the statistical phase of the above procedure is considered, researchers face the problem of selecting a technique. The measurement level of the respective variables determines this choice:

a. Where T, X and Y are quantitative, partial correlations can be used. As long as the researcher is cautious, dichotomized qualitative variables can also be considered quantitative.

b. Where T, X, and Y are qualitative, partial gammas can be used. Here analysis is simplest if all the variables are dichotomous.

c. Where Y is quantitative and X and T are qualitative, two-way analysis of variance can be used.

d. If Y is qualitative and either or both T and E are quantitative, all 3 variables will have to be made qualitative. Then partial gammas can be used.

When these data are interpreted and reported, it is much easier if all the 3-way computations use an identical technique--that is, only partial correlations, only gammas, or only analysis of variance (ANOVA). Provided that all the steps outlined earlier have been followed (cf. Preparing the Data section above), partial correlations will require the least computation* and produce statistics that are at least as easy to interpret as the other techniques.

The statistics that can be obtained from the respective techniques are:

	DESCRIPTIVE	INFERENTIAL
Partial Correlation	Partial r	Confidence Interval or Level of Significance
Gamma	Partial Gamma	Through computation of Chi-2 can
Analysis of Variance	Means and Adjusted Means	F-score and level of significance for contribution of E.

*

Partial correlation coefficients can be computed from simple correlation coefficients, but with Gamma or ANOVA the

The statistics from analysis of variance are usually the most difficult to understand and compute. By contrast, partial correlation statistics are the easiest. To summarize the above discussion, 3-way analysis is useful when the researcher wants to show that a key relation between two variables is valid, independent of the influence of a select group of other relevant variables. The task is easier when all variables are measured at either the quantitative or the qualitative level.

Three-Variable Analysis in Causal Models:

As long as all three variables share a common measurement level (quantitative or qualitative) models of the causal interrelations of the variables can be postulated, and the fit of the data to these models can be tested. Blalock (1964) showed how this could be done when the variables were measured at the quantitative level, and Davis (1971) for the qualitative level. Causal modeling with three variables is useful for pedagogical purposes, but is rarly relevant in policy analysis. There are few situations where there are only three variables of interest. In most situations, the initial focus on three variables will expand to consider others. Hence, the researcher will need to turn to the more powerful set of basic and derived multivariate techniques described later.

The Utility of Partials:

While the discussion above focuses only on three variables, the formulas for both partial correlations and partial gammas can be extended to control for 2 or more T-variables simultaneously. This extension provides an even more rigorous test of the strength of the X-Y relation. However, the computation is burdensome and best performed with the aid of a computer.

In carrying out this extension, the researcher must carefully select the variables for the control situation. In general, only those variables that are thought to causally and temporally precede Y should be selected. If only one T variable is selected, the result is a first-order partial; if 2, a second-order partial, and so on. The results of such a partial analysis are extremely useful for researchers if and when they decide to undertake a more complicated multivariate analysis. The variables that continue to have strong partials, even after being controlled by several other variables, are the

strongest candidates for inclusion in a more advanced multivariate analysis. The statistics for second and higher order partials are identical with those for first-order partials.

Now that relational multivariate techniques have been discussed, the next section will describe the most common multivariate techniques.

2. Basic Multivariate Techniques

This section presents the three most common types of multivariate techniques: (A) Multiple Regression, (B) Analysis of Covariance, and (C) N-way Analysis of Variance.

A. Multiple Regression:

Multiple regression is a remarkably flexible technique not only in terms of the measurement level of variables it can handle and the goals it can tackle, but also in terms of the computational procedures employed. Essentially, there are three computational forms. Researchers have to be aware of the differences in them and in their use.

Simple Multiple Regression: The best known form is called simple multiple regression, where the dependent variable is regressed on a selected group of independent variables, all at once. The result is an equation, expressed in unstandardized partial regression coefficients (b's). This equation is useful to make predictions. When expressed in standardized partial regression coefficients, the equation is useful to identify the variables that are most important. Simple multiple regression can accomplish most of the analytical goals discussed at the beginning of the chapter.

Simple Forward Stepwise Multiple Regression: Sometimes the researcher feels there are too many variables. Consequently, only a few of the most powerful seem necessary for the equation. In such a case, the researcher might select simple forward stepwise multiple regression, a procedure which adds the most powerful independent variable, the next most powerful second variable, and on down the line until the new variables add very little to the total explained variance in Y. The reverse of the above form is simple backward multiple regression, a procedure less commonly employed.

Hierarchical Multiple Regression: Usually, by the time a researcher turns to multiple regression, the variables have been sorted into sequential groups or blocks starting with the most independent and moving towards the most dependent. For some problems, it is desirable to consider the impact of each of these blocks in sequence, starting with the most independent block and moving forward. If so, the researcher elects a hierarchical multiple regression. As with simple forward stepwise regression, a stepwise procedure is used. But unlike simple regression, the researcher specifies the order in which the respective blocks of variables should be entered. In the extreme case, each block might contain only one variable, but it is more common for a block to contain several variables which all share a common temporal, conceptual, or logical character.

Within blocks, it is most common to have each variable enter according to the simple stepwise form. In hierarchical multiple regression, the main statistic of interest is the amount of variance added by each successive variable and by each successive block.

While each of these procedures has different assumptions and uses, the computer is "dumb". Thus, the computer is likely to pour out the same mass of statistics regardless of the form one is following.

Table 4.2 on the next page indicates with an "x" which statistics are most likely to be of interest for the three forms of multiple regression.

Table 4.2

Statistics Used for the Forms of Multiple Regression

Statistics Produced by Computer	Simple Regression	Simple Stepwise Regression	Hierarchical Regression
Simple r	x		
Partial R Controlling for all variables in Earlier Group.			x
Unstandardized Partial Regression Coefficients	x	x	
Standardized Partial Regression Coefficients		x	
Proportion of Explained Variance added by Each variance.		x	x
Total Explained Variance	x		
F-score and Standard Error for b's each Variable's b	x	x	
F-score and Standard Error for Multiple Correlation	x	x	

Multiple regression is the most frequently used multivariate technique. Moreover, it is the most efficiently computed. Unlike the other techniques, multiple regression is considerably valuable to describe, explain, and predict. The standard outcome of multiple regression is an equation that makes Y a function of several Xs as follows:

$$Y = a = b1\ X1... + b2\ X2... + bi\ Xi$$

If the values in the regression equation adequately summarize the major factors that determine Y, the researcher can make good use of this equation to describe the content of Y. For example, if Y is family income in thousands of rupiah, and the Xs include education of father, education of mother, number of children, distance of home from center of city, numbers of hectares of land owned, and so on, the researcher can make precise statements about how much change in each of these independent variables would be required to increase income by 10,000 rupiah. Possibly, decreasing family size by .2 children or raising father's education by 2 years would achieve the same results. In cases where 2 of the dependent variables are manipulable variables, conclusions about which manipulation is most practical and will have the greatest impact will follow. None of the other basic techniques has this descriptive feature.

The regression equation is also useful to predict. If researchers have collected data from an appropriate sample, they can infer that the values of the a's and b's obtained will be generalizable to other samples. Thus, when given a set of X values for a new case, these values can be inserted into the earlier equation to predict the Y for the new case. Predictions like this are often useful for planning exercises.

The b's in the equation are a function of both the relative importance of the X's they modify and the range of the metric used to measure these Xs. By simply standardizing the metrics of all of the variables in a particular analysis, the researcher ends up with a new set of statistics, Betas, which precisely reflect the relative contribution of each of the Xs to the equation. These Betas are immensely valuable for explanations. They can be thought of as analogues to correlation coefficients or partial correlation coefficients, although they are different.

As long as categorical variables are entered as dummy variables, multiple regression can be used as a substitute for analysis of variance. Discriminant analysis, at least when the dependent variable is a dichotomy, is identical with multiple regression. Thus multiple regression is the most general technique.

B. Analysis of Covariance:

This technique is very close to regression. Analysis of covariance, rather than entering certain independent variables in competition with others to explain the variance in the dependent variables, allows these variables to enter first and take away their share of the dependent variable's variance. The result is a new dependent variable controlled for these covariates. The new dependent variable is then regressed on the remaining independent variables to highlight their special influence.
Usually these remaining independent variables are qualitative, but they need not be exclusively qualitative.

C. N-Way Analysis of Variance:

Of the several techniques mentioned in Table 4.1, N-way Analysis of Variance has the weakest relation to multiple regression. The standard use of N-way analysis of variance is to determine whether 2 or more groups produce a different outcome (measured with a quantitative metric), controlling for the influence of other qualitative variables. While multiple regression assumes that each independent variable has a linear relation to the dependent variable, analysis of variance simply assumes that the categories of the independent variables have different mean scores on the dependent variable. Inferences can be made about the relative importance of the independent variables, but these are somewhat clumsy. There is no analogue in the usual N-way analysis of variance to the b's or Betas of Regression Analysis.

The third group of multivariate techniques to be described are the derived techniques of multiple classification analysis and commonality and path analysis.

3. Derived Multivariate Techniques

This section will cover the derived multivariate strategies: (A) Multiple Classification Analysis, (B) Commonality Analysis and Path Analysis.

A. Multiple Classification Analysis

This technique is derived from analysis of variance to produce these analogues. In N-way Analysis of Variance, the procedure is to decompose the mean scores on Y for each of the categories of each of the independent variables into those components that come:

- From that category,
- From other independent variables,
- From unmeasured sources.

Standard analysis of variance output does not report this decomposition, but is only concerned with showing that the decomposed or adjusted means of the categories of each independent variable are significantly different from each other. However, in multiple classification analysis, both the unadjusted and adjusted means are reported. Moreover, a Beta is computed which is a rough analogue to the Beta of regression analysis. This Beta can be used to make inferences about the relative importance of each independent variable. The additional information produced by Multiple Classification Analysis usually significantly enhances the research team's ability to understand and explain the findings.

B. Commonality and Path Analysis

Apart from the techniques discussed above, commonality and path analysis, derived from simpler techniques, are useful in causal explanations. Commonality analysis is derived from simple Pearson correlation coefficients. Although path analysis can be derived from simple correlation coefficients, it is more convenient to use the Betas from multiple regression. Thus both of these techniques require the component data to be measured at the quantitative level.

It would be most unusual for a researcher, especially when an applied problem is the focus, to plan to use one of these techniques from the start. However, once the

researcher completes the analysis with basic multivariate techniques, further understanding as well as more lucid presentation can be possible, by using one or the other of these techniques.

Commonality analysis is useful in a situation where several (not too many) independent variables that have an almost identical temporal location are known to jointly influence a dependent variable. Commonality analysis can sort out the joint and unique effects of each of these.

Path analysis is more useful when the independent variables can be ordered in a series of causal-temporal steps from the most to the least appropriate. Calculations of path coefficients illustrate and enable comparison of the direct effects of the more approximate variables relative to the more distant variables. Additional calculations can reveal the total effects of all the component variables.

These techniques--while quite popular--can easily be misused. The researcher should be especially careful to seek the informal advice and criticism of colleagues before publishing results in a policy report.

THE SIMILARITIES OF ALL MULTIVARIATE TECHNIQUES

The various multivariate techniques just described have been independently developed by researchers to solve the seemingly unique statistical problems over the past several decades. Only in the past decade have researchers come to realize that most of the established techniques have fundamental similarities. Hence, they should be considered as a set of interchangeable tools, rather than as a set of distinctive tools for unique purposes. This new awareness was perhaps best stated by Fennesey in his outline of the General Linear Model. Given the similarity of all these techniques, it is still useful to distinguish between the basic teachniques of multiple regression, analysis of variance, and log-linear models on the one hand, and other techniques, which are simply derived from the standard statistical outcomes of these basic techniques. There are striking similarities in the basic techniques which deserve mention:

1. The common intent of all the "basic" techniques is to show the interrelation of a single dependent variable with 2 or more independent variables. In each case, independent variables are added in an effort to explain the variability of the dependent variable, or reduce the unexplained variance. Thus one outcome of all these techniques is a statistic that indicates the total amount of variance explained. For multiple regression and discriminant analysis, the statistic is R2, whereas for analysis of variance and covariance, it is Omega 2.

2. The number of independent variables that can usefully be included in these analyses is always limited. Two general criteria are helpful when selecting independent variables:

o The independent variable should have a <u>meaningful</u> relation to the dependent variable.

o The independent variable should be known or presumed to have a <u>strong</u> relation to the dependent variable.

In addition, where the goal is simply to reduce unexplained variance, it is best to choose independent variables that have no close associations with each other. A useful criterion for evaluating a particular variable that may be added to a multivariate analysis is to see whether it adds much to the amount of variance already explained. Depending on the size of the sample and other factors, only 2 to 5 independent variables are likely to make a substantial contribution.

3. The F-score is the statistic normally used to evaluate the probability that the amount of variance explained in the dependent variable for the sample under study is likely to be similar for the population. The F-score comes from calculations of the explained sum of squares divided by the total sum of squares. While the procedures for F-score computation differ according to the technique, the basic logic is the same for all the techniques described above. Likewise, the calculations for determining degrees of freedom are similar.

4. The F-score is the common statistic used to evaluate the probability that the contribution of each of the independent variables is not by chance.

Naturally, differences in the techniques also deserve mention. Many of the differences were noted as the three groups of techniques were described in the previous sections. The most important difference illustrated in Table 4.1, though, is in the measurement level of the variables they can handle. None of the above techniques is well-suited for situations where both dependent and independent variables are measured qualitatively. The basic approach in these situations is either to produce complex sets of cross-tabulations or to alter variables so they are compatible with the more standard techniques. Over the last decade, new techniques, such as log-linear models have been developed for analyses in these situations, but such techniques are not yet feasible, and moreover, are very expensive to use. Fortunately, few of the problems currently researched fall into these groups.

The final two sections of this chapter will discuss planning and reporting a multivariate analysis.

PLANNING A MULTIVARIATE ANALYSIS

This chapter began by observing that most applied educational research cannot be effectively completed without multivariate analysis. This is because the educational system is composed of many variables, with complex patterns of relations linking them. While multivariate analysis is required, the initial goal is usually relatively elementary, that is, to simply determine which of these relations is truly significant. However, once researchers solve the basic problem, they may become interested in a more challenging goal. They want an approach that efficiently allows flexibility. Finally, researchers may be limited both by their training and ability. Therefore, they need a technique that has the most general applicability.

It is advisable that the researcher, in most cases, should follow a standard set of procedures:

1. Order the central concepts and their indicators from those most dependent to those most independent. The indicators can be broken down into 2 or more groups, depending on their temporal and logical characteristics. The most dependent should belong to the Y group, the next most so to X, then T S, R, and so on. Usually 3 or 4 groups are sufficient.

2. The data should be cleaned, and basic univariate descriptive statistics computed.

3. All measurement levels should be reviewed. Where practical, when a variable has been measured at a qualitative level, new dichotomous variables, dummy variables, or quasi-interval scale variables should be created from it.

4. Bivariate correlations between all the variables should be computed and saved in a correlation matrix.

5. For all X-Y relations of interest, partial correlations should be computed controlling at least for all variables of the T group and all earlier (i.e. more independent) groups. When it seems appropriate, even some of the other X's can be included as controls in these partial correlations.

6. When appropriate, the researcher should move into the family of basic multivariate techniques. Most likely, multiple regression will be the preferred technique given its wide variety of applications and the relative ease of computation. Given the wide applicability of multiple regression, the researcher will have to concentrate on planning an analysis that matches the goal.

7. Under exceptional circumstances, the researcher may decide to move on to the derived techniques appropriate for causal explanations.

8. Once the analysis is complete, the researcher must select an appropriate style for reporting results.

REPORTING A MULTIVARIATE ANALYSIS

Multivariate analysis is immensely important, if the researcher wishes to produce believable findings. But even if analysis is done correctly, there is no guarantee that the findings will be believed. The main obstacle is that few policy makers are familiar with multivariate analysis. Unless a report is able to effectively communicate the complex and extensive finding simply and comprehensibly, the research will have been futile.

One general rule that should be followed in reporting multivariate analysis is to keep the language simple and to put many of the statistics in technical appendices. Eight other suggestions are provided starting on the next page.

1. Early in the report, the researcher will already have discussed all of the variables in the study. In a report of multivariate analysis, the decisions on including and excluding certain variables should be considered. In a footnote or appendix, the researcher should also report the transformations or other manipulations engaged in to enable the data to conform to the assumptions of the technique used.

2. The researcher should review in some detail what are the likely outcomes of the analysis, and which of these are most promising for the policy maker. This discussion should be logical, with many explanations.

3. The method to be employed should be reviewed briefly.

4. To help the reader, the report might briefly summarize some other instances where the same method has been used, showing how it helped to reveal interesting findings in those cases.

5. While the researcher should compute simple associational measures between all of the relevant variables as well as partial correlations, these should not be reported in the body of the report. They are, however, most appropriate for an appendix, because they facilitate replication of the work.

6. Following a discussion of methods and expectations, the report should turn to an initial finding and review it in some detail. It is undesirable to report <u>all</u> of the statistics associated with the result. Nevertheless, statistics that are relevant to the argument should be explained and interpreted in detail.

7. Subsequent findings can be reported, especially using tables with verbal discussion restricted to highlighting the most interesting and relevant points. Where there are many findings, the researcher may be selective, and concentrate only on those that do have implications.

8. If the analysis progresses into causal models, a separate chapter might be prepared for them. The researcher should always remember to try to help readers through the necessarily complex arguments presented.

WEEK 3: EXERCISES

OVERVIEW

These examples and practice sessions are designed to supplement the content of the Workshop's third week. They provide examples of the kinds of questions dealt with in policy research and computer potential to make research tasks easier. The examples will deal largely with descriptive issues. Although the examples may only prove the obvious, nothing should be taken for granted in research. By examining relationships that appear evident, our goal is to provide clear examples of the computer's usefulness in research.

POLICY ISSUE

In Sri Lanka, the management reforms encouraged by the Ministry of Education in 1984 intended to improve the quality of education and the effectiveness of less-advantaged schools in rural regions of the country. The reforms aimed at reducing the disparities in quality between better schools and the less advantaged schools. To understand the context in which these reforms were implemented, as well as to assess their impact, independent of any differences in the affluence of their constituencies, the relative socio-economic status of the various districts and the differences between rural and urban areas must be examined.

Using data from two survey instruments administered to the principals of schools in the study's sample, we will build an indicator of socio-economic status and examine its distribution in relation to rural/urban location and to the type of school (primary only, O-level preparation, or A-level preparation). Examples will feature both LOTUS and SYSTAT software packages.

To improve the quality and effectiveness of education in less advantaged schools, the reforms aimed at promoting increased community contributions to the school. Through additional statistical methods, we will examine the relationship between reform implementation and level of community contributions to the schools, controlling for the socio-economic status of the community. Examples will use the SYSTAT software package.

EXAMPLE 1 : CREATING AN INDICATOR and UNIVARIATE ANALYSIS

The first step in our analysis example is to create an indicator of socio-economic status (SES). In the sociological literature, the major constituents of SES are income and education. Although we have no information from individual households of families that send children to school, we do have relative estimates by school principals. We will construct a composite variable with the name SES using data gathered in Principal's Questionnaire items 4.3 and 4.4. Refer to excerpts from the Principal's Questionnaire for the content and coding of the items.

A. Locate the necessary variables in the Rbase data file, and read them into a LOTUS spreadsheet.

1. Note that the variable names begin with Q. This indicates that the question is drawn from the Principal's Questionnaire.

2. Questions drawn from the Principal's Survey are coded to begin with an S.

3. The variables you want are named Q041 and Q042.

B. Look at the codebook and decide how the variables as coded relate to each other.

Note that the question on economic level of parents (variable Q043) has possible values from 1 to 5:

1 indicates a 'very good' economic level
5 indicates a 'poor' economic level.

By contrast, the question regarding parent's education (variable Q044), which also has possible values from 1 to 5, is coded differently. The direction of the scale is opposite that of Q043:

1 indicates the lowest level of education (did not complete primary cycle)
5 represents the highest level (received a higher education).

In an actual project, we would want to find out how these variables are distributed, and how they relate to each other statistically. Since this is a sample exercise, we will wait till later to describe them using univariate statistics.

We will not calculate correlation. Variables Q043 and Q044 are negatively correlated with each other as originally coded.

To create the composite indicator of SES, we will add the values to combine the information in them. However, in their present form, adding the values would cancel the effects. For example, if a principal reported that the majority of student parents at the school were 'very well-off' economically (a score of 1 on Q043) and had 'received higher educations' (a score of 5 on Q044), simply adding the reported values would give a composite SES score of 6. The same SES score would be computed for a school where parents' economic level and education were reported as the middle values. If education and economic condition vary in the same direction, combining the scores in their present forms would erase differences between schools, rather than elaborate differences. Thus, before creating our composite measure, one additional step is necessary. The next section C presents the steps to take, in order to scale values in the same direction.

C. Recode the variable Q043 so that a higher economic level is coded with a higher number. This will result in both components being scaled in the same direction.

1. In the column next to the two variables you have used, create a recoded version of Q043. In order to keep track of your data, preserve the basic form of the variable name. You might choose to call it XQ043, NQ043, of whatever will help you identify your operations.

2. Put the name of the new variable at the top of the column.

3. Enter the formula to calculate the new variable in the second row of the column. The formula to reverse the scaling is 6-Q043. Thus, a value of 5 becomes 1, while a value of 1 becomes 5.

4. Copy this formula into the rest of the necessary cells (through row 274). Now you are ready to create the composite indicator of SES.

D. Create the composite indicator of SES in the column next to the recoded version of Q043.

1. Enter the name of the new variable (SES) at the top of the column.

2. Enter the formula to add the new version of Q043 to the value of Q044 in the second cell of the column.

3. Copy this formula into the remaining 272 cases in the column (again, through row 274).

You have created a new variable that captures the significant aspects of the socio-economic status of the parents of a school's students.

4. Create a composite indicator of socio-economic status called SES.

E. Examine the frequency distribution of the SES variable.

Create a frequency distribution table for the new variable SES, using the Data function of LOTUS. Read all instructions before proceeding.

1. Select Data from the command line at the top of the screen.

2. Select Distribution from the /Data command line.

3. The prompt will ask first for the 'Value Range'.

4. Highlight the range containing the values for that variable, because we are interested in the distribution of SES. Do NOT include the first row in the column, which contains the variable name).

5. The prompt will ask you for the 'Bin Range'. This range contains the possible values for the variable. LOTUS will enter the number of observations at each of the values in the 'Bin Range' to the right of the number. It will enter the number of observed values that do not fit any of the categories below the last number in the column.

6. Be sure that the cells to the right of the Bin range and the cell beneath and to the right of the last cell in that range are open.

7. Enter the possible values in a 'Bin Range'.

Since we created SES by combining the recoded version of Q043 and Q044, both of which had possible ranges from 1 to 5, our new variable has a possible range of 2 to 10. No school should have a value lower than 2, none can have a value higher than 10.

8. Place some descriptive title in the first row of a column to the right of the columns you have been using.

9.Enter the values 2 through 10 in ascending order beneath that title.

10. Proceed with the command sequence outlined above. This will produce a frequency table for the variable SES.

11. Print this range for future reference, for the next examples.

F. Perform additional Univariate Analysis on the variable SES and its constituent and source variables Q043, Q044, and the recoded version of Q043.

LOTUS allows you to easily describe the distribution of the variable.

1. Move 2 or 3 columns to the right of the frequency distribution table you have created.

2. Enter the heading UNIVARIATE ANALYSIS in the top row. Beneath this heading, leave one blank row and enter the following abbreviations in a column:

AVG	(the mean of the values in the distribution)
COUNT	(the number of non-blank entries)
MAX	(the maximum value observed)
MIN	(the minimum value observed)
STD	(the standard deviation of the distribution)

In the blank row beginning with the cell to the right of the names of the variables SES, Q043, Q044, NQ043, your table should look like this:

UNIVARIATE ANALYSIS

	SES	Q043	Q044	NQ043
AVG				
COUNT				
MAX				
MIN				
STD				

In the cells directly to the right of the labels you just entered, enter the formulas listed below with appropriate ranges for each variable.

@AVG(range of the SES values)
@COUNT(range of the SES values)
@MAX(range of the SES values)
@MIN(range of the SES values)
@STD(range of the SES values)

LOTUS will calculate the appropriate statistics and enter them in the cells.

G. Print out your findings.

1. Select the range containing the frequency distribution and the table of univariate statistics.

2. Print.

3. Save this printout for future reference.

REMEMBER TO SAVE YOUR FILES

EXAMPLE 2 : BIVARIATE ANALYSIS

You have now created an indicator of the socioeconomic status of our sample school constituencies. You have also performed univariate analysis on the variable, and found that it varies sufficiently to discriminate among schools. If all the schools had a value of 6 on SES, the variable would only suggest that the schools are all the same in that regard, and that no further analysis of that aspect of the school's environment is needed.

In this example, we will use the variable SES to describe differences among the schools in our sample along dimensions, comparing schools grouped by district and by type.

A. Find the variables SCNO, C86(4), and C86(8) in the Rbase files, pull them into a LOTUS worksheet, and add the variable SES which you created in the last example.

1. Arrange the variables so the variable name appears in the top row, and the values begin in the cell immediately below the name. The variable SCNO is the school's number.

2. When performing analysis that requires data sorting, it is essential to include such a marker variable in the data file. By looking at the SCNO column in a data file, you can see whether the data has been sorted or is in its original order.

B. Use the /Data capabilities of LOTUS to sort your data by various criteria to analyze differences between groups of schools.

Sorting the data is a preliminary step to examining differences between groups of schools. LOTUS allows you to sort your data on one or two variables. We will initially sort the data by district (variable C86(4)).

1. From the main command line, select **Data**.
2. From the /Data command line, select **Sort**.
3. From the /Data Sort command line, select **Data Range**.

LOTUS will ask you for the range of data you want to sort.

4. Select the whole data range.

DO NOT INCLUDE THE TITLES IN THE FIRST ROW IN THE DATA RANGE.

5. Select rows 2 through 274 for each column full of data. (Note: You can select the range either by typing -- for example, a2..d274 -- or by 'pointing' to highlight the range.)

6. Once you have selected the data range, the /Data Sort command line returns.

7. Select **Primary Key**

8. Enter the address of the first data cell in the column containing the variable which codes district.

9. LOTUS will then ask whether you want it sorted in Ascending or Descending order.

10. Type the letter **a** to indicate ascending. (This means LOTUS will rearrange the rows of data with all those rows with the lowest value for C86(4) at the top, and so on.)

11. When the /Data Sort command line returns, choose the option Secondary Key. You will now sort the data by a second variable, C86(8), which codes the school's type (see the codebook for an explanation of this variables' coding).

Repeat the steps used in defining the primary key above, selecting C86(8) as the secondary key. LOTUS will sort the schools in each district by their type.

When the /Data Sort command line returns, select **Go**. LOTUS will then sort the whole data range.

C. Create a table showing the mean SES of the schools of various types in the various districts.

1. To the right of the data columns, enter titles for your table as in the example below.

2. Enter the formula used previously to calculate the mean -- @AVG(range). The range entered for each line will indicate the rows in the SES column that contain the records of schools in a particular district of a particular type, as indicated in the columns containing the variables C86(4) and C86(8).

MEAN SES BY TYPE AND DISTRICT

DISTRICT	TYPE Ia&b	TYPE Ic	TYPE II	TYPE III
1	@AVG(..)	@AVG(..)		
11				
17				
27				
28				
31				

3. To fill in the appropriate ranges, look down the column of data to see which rows pertain to each district and each type. Use the **Worksheet Titles** command.

4. Position the cursor at a cell in row two, and execute the command sequence **Worksheet Titles Horizontal**.

This will keep the titles in row 1 visible as you scroll through the data to complete the formulas.

Note that the first 20 rows of the data do not register values for the variable C86(4). These are the private schools in the sample which are not part of the MOE's official district educational structure.

D. Print out the tables you have produced.

EXAMPLE 3 : INTRODUCTION TO ANALYSIS

USING THE 'SYSTAT' STATISTICS PACKAGE

In this example, you will use the SYSTAT statistical package to repeat the analysis you completed in Examples 1 and 2. While spreadsheet packages such as LOTUS are powerful tools for data analysis, statistics packages such as SYSTAT, SAS, SPSS, and others, can greatly ease the work of analysis and allow the researcher to perform more complex analyses.

The analysis of socio-economic status in school constituencies is repeated, using SYSTAT and Rbase.

A. Enter the DATA module of SYSTAT and bring together the following variables; SCNO, C86(4), C86(8), Q041, Q043, and Q044.

Two different procedures can be used to call up data, depending on whether you are using SYSTAT or RBASE. First we present the steps with SYSTAT, followed by the steps in RBASE.

SYSTAT

1. The SES variables are in two separate files: CENDATA.SYS, and PRQDATA.SYS. Enter the following command at the prompt to call them up:

>use cendata(scno c86(4) c86(8)) prqdata(q041 q043 Q044)

2. SYSTAT will respond by listing the variables available to you. Note that SYSTAT does not display the data as does a spreadsheet program. Rather, it holds the data in its memory, waiting for you to enter commands.

RBASE

1. Create an export file, called EXDATA containing those variables in the delimited ASCII format.
2. Change the file extension in the last 3 characters after the . full stop to ".DAT"
3. Use the command:
   > get exdata
4. Retrieve data.

B. Prepare the variable we will use in analysis by recreating the SES variable.

1. Enter the following command to tell SYSTAT to create the variable SES for each case in the file. (Note that you do not need to copy the formula into a cell for each case as you did in LOTUS):

```
>let ses=q044+(6-Q043)
```

2. When you hit ENTER, SYSTAT will check to make sure it understands your command. **SYSTAT will not perform the calculation yet.** SYSTAT performs some commands instantly (as your manual explains, these are its !HOT! commands), but most are held in memory until you give it the **RUN** command.

3. One additional note regarding SYSTAT's calculation: if a value is missing for any of the variables used in a calculation, SYSTAT considers the value of the equation to be missing. For example, if a school's record is missing a value for either or both of Q043 or Q044, SYSTAT will give that school a missing value (coded as a period, ".") for SES. This avoids the problem encountered in LOTUS which assumes a zero for blank cells.

4. Indicate where you want the data on hand saved, before telling it to run. Enter the following command:

```
>save exampl3a
```

When you tell SYSTAT to run the short program you have given, it will save the processed data in a file called EXAMPLE3.SYS.

5. Enter the command **RUN**. If you receive an error message after entering the command to run, ask one of the lab assistants for help. SYSTAT will always report to you how many variables and how many cases were processed.

6. If you want to see if the formula really worked, look at the data using the list command in the Data module.

Call up the file you just used (type **use exampl3a**)
Type the following commands:

```
>repeat=20
>list Q043 Q044 SES
>run
```

SYSTAT will then list out the first 20 cases of the variables you specified in the command. (Note that you could have told SYSTAT to print the output by including the command output @ on a line before the **run** command.

C. Perform univariate analysis on the variables in the data file using the STATISTICS module.

1. Leave the Data module by entering the command QUIT. This will return you to the opening menu.

2. Enter the number for the Statistics module. This module allows the researcher to perform analysis of distributions.

3. Type HELP for a list of the commands available to you in this module.

To obtain a summary of the distribution of each variable, we will use the command STATISTICS. Since this is a 'HOT' command, however, we must first enter a command to route the output to the printer so we have a hard copy of the results.

4. Send output to the printer by entering this command:

```
>output @
```

5. Enter the command word followed by the variable names for which analysis is desired.

```
>statistics Q043 Q044 SES
```

Note that we did not ask for statistics for C86(4) or C86(8) because these variables are not 'continuous', or 'quantitative'. We will use another module to examine their frequency distribution later.

As soon as you entered the last command, SYSTAT began calculating the statistics for those three variables. It will list out the MAXIMUM, MINIMUM, MEAN, and STANDARD DEViation for each. Note that it also lists the number of observations for each. The fact that not all had the same number of observations shows that there was missing data in each variable.

D. **To see what SYSTAT can do, check the relationship between the constituent parts of SES, Q043 and Q044.**

1. Leave the Statistics module by typing **quit** <Return> at the prompt.
2. Choose the number for the Correlation module.
3. Call up the same file, Exampl3a.sys.
4. Route the output to the printer
5. Enter the command:

>pearson q043 q044/pairwise

This command tells SYSTAT to produce a Pearson correlation matrix for the two variables. Their relationship should be strong and negative (remember, we had to recode Q043 to change its direction). The last part of the command, **/pairwise**, instructed SYSTAT to exclude cases with missing values on either variable from the calculations.

E. Examine the 'categorical' or 'qualitative' variables using the TABLES module.

1. Type QUIT to leave the STATISTICS module. SYSTAT will produce a list of the commands used in the module and include them in the print-out it just generated. (Note: If the printer has been turned off or taken 'off-line', it will not continue. Check the printer if there's a problem).

2. Enter the TABLES module from the main menu.

3. Call up the file EXAMPLE3A.SYS once again.

4. Before requesting analysis, tell SYSTAT where you want the output sent.

5. Enter **output @** as in the previous step.

6. To generate a description of the distribution of the categorical variables C86(4) and C86(8), enter this command:

>tabulate c86(4) c86(8)/list

SYSTAT will now produce a printout describing the frequency distribution of those variables, telling you how many schools are in each district, and how many are of each type. This is the same information that required you to enter the data range and the list of possible values in Example 1 in LOTUS.

7. To find out how many schools of each type are in each district, enter the command:

```
>tabulate c86(4)*c86(8)
```

Note that this produces an accurate count of the groupings of schools without the necessity of scrolling through the data on the screen as you have done in the spreadsheet.

F. Prepare the data file for bivariate analysis by sorting and resaving it.

In order to perform statistical analysis on groups of schools, SYSTAT requires that the data file be sorted, just as in LOTUS. But in SYSTAT, sorting is much simpler.

1. Re-enter the Data module
2. Call up the data file by entering the command:

```
>use exampl3a
```

SYSTAT will now list the variables available to you. The list should now include the SES variable which you recreated. Our next task is to sort the file, just as we did in LOTUS. The task is easier in SYSTAT.

3. Enter the following command:

```
>sort c86(4) C86(8)
```

This tells SYSTAT to sort the data records according to two key variables, first by district (variable C86(4)), then within districts, by type of schcol (variable C86(8)). Again, SYSTAT will check your command to ensure that it has the correct variables on hand.

SYSTAT will not run the command until you enter a **save command** and a **run command.**

4. Save this file with the name EXAMPL3B.SYS. (Note: It is important to keep track of your file names so that you know what each file contains and how it has been sorted or otherwise manipulated.)

G. Now leave the Data module, re-enter the Statistics module, and analyze the variable SES by District (variable C86(4)) and by type of school (variable C86(8)).

1. Tell SYSTAT to send the output to the printer. (The command is **output @**)

2. Tell SYSTAT that your analysis will be of groups of schools, not simply the entire column of data. Enter the following command:

```
>by c86(4) c86(8)
```

You must enter the 'by' variables in the same order that you entered them in the sort command, or SYSTAT will not be able to complete the analysis if the order is reversed.

3. Enter the command to produce statistics for the desired variables. To reproduce the table we want, we only need to look at SES. In order to compare it with its constituent parts, Q043 and Q044, we can ask for all three. Enter the following:

```
>statistics ses q043 q044
```

SYSTAT will now send to the printer a separate analysis of the SES, parent's education level, and parent's economic level for each type of school in each district. The printout will contain the same information for each group as in the previous summary printout; N, MIN, MAX, MEAN, and STD.

4. From this printout, produce a table using paper and pencil, LOTUS, or a typewriter, to show the mean SES of each group of schools.

5. Save your printout and your table for future reference and later discussion.

EXAMPLE 4 : TRANSFORMING VARIABLES

AND ADDITIONAL ANALYSIS USING SYSTAT

As the section on multivariate analysis indicates in Chapter 4 of your workshop manual, careful attention must be paid to the level of variable measurement, whether qualitative or quantitative, when you plan analysis. In this example, you will transform two variables for use in later analysis.

The qualitative variable Q041, which assigns each school to a category of urban, suburban, or rural location, will be transformed into a dichotomous variable (as described in the manual, Chapter 4). The continuous or quantitative variables S463a, S463b, and S463c, which record the value of monetary, in-kind, and labour contributions of the community to the school, will be divided into categories. In performing this transformation, the underlying quantitative nature of the distribution will not be changed. The schools will merely be grouped by the size of the contributions received.

A. Call up the SYSTAT file EXAMPL3A and the variables S463a, S463b, S463c.

1. Enter the Data module of SYSTAT.

2. YOU MUST USE THE UNSORTED VERSION of your previous data file.

3. Merge the data together with the following command:

```
>use exampl3a prsdata(s463a s463b s463c)
```

Our goal is to transform some of these variables for use in further analysis. But as the workshop manual points out, the data's distribution has to be carefully examined before such variables can be chosen. Thus, before we begin to transform the variables, we will save the file and examine the data in their present form.

4. Save this file under the name EXAMPLE4A.

B. Produce descriptions of the distributions of the variables.

1. Enter the Tables module in SYSTAT
2. Call up the file you have just created.
3. Use the **TABULATE** command to get a printout of the frequency distribution of the variables.
4. As before, route the output to the printer.
5. Enter the output command.
6. Enter the following command:

```
>tabulate q041 s463a s463b s463c /list
```

The /list portion of the command tell SYSTAT in which of several optional formats you want the data displayed.

7. Enter the Statistics module.
8. Route the output to the printer
9. Get descriptive statistics for the contributions variables. Use the command:

```
>statistics s463a s463b s463c
```

10. Look over these descriptions of the data. As for the distribution of Q041, note that 40 cases are coded as missing.

Next, we will create a variable called RURAL which will indicate whether or not a school is in a rural location. Combining the suburban and urban schools in the non-rural category will provide a larger group for comparison (note that the decision to combine these categories depends more on the realities of similarities and differences in the country and on prospective analytic techniques, NOT on the desirability of group sizes.)

Also notice the numbers of observations coded as missing in the contributions variables. As researchers, you will make a mental note that analysis of these variables, given the number of missing cases, will impose difficult choices in coding, transforming, and combining to produce an overall community contributions index. In our next step we will retrace the steps of the researchers who produced a report from these data in categorizing the contributions.

C. Practice the creation of a dichotomous variable from the qualitative variable Q041.

1. Return to the Data module of SYSTAT.

2. Call up the file EXAMPL4A and the additional variables E86(35) and E86(5\36) from the CENDATA.SYS file.

You can enter the Save command at any point when you are working in the Data module. This time, we will enter it first. This strategy ensures that we will not forget where we want the data stored in its new form.

3. Enter the Save command for the filename EXAMPL4B.SYS. (Remember, the identifying suffix '.SYS' is added automatically by SYSTAT -- you do not have to type it).

4. Create a variable called RURAL from the variable Q041. As you see in the codebook, the value of Q041 is 1 for rural areas, 2 for suburban areas, and 3 for urban areas. Our new variable RURAL will take a value of 0 for non-rural areas, and a value of 1 for rural areas.

In creating a new variable of this type, be careful not to generate false data. For example, one strategy to produce the new variable would be to set the value of RURAL to 0, then change the value to 1 for those cases where Q041 indicates a rural school. However, as our frequency distribution indicates, 40 schools are coded as missing. Such a strategy would give these schools values of 0 -- indicating that they were non-rural schools when in fact this may not be the case. **Missing data should not be replaced with values unless the researcher is absolutely certain of the validity of the data created.**

If we examine the frequency distribution, we are alerted to another problem in the data. Two schools have values of 4 in Q041. We have no way of knowing what these values mean. In this case, it is safest to set the values to 'missing', which is coded as '.' in SYSTAT.

Now we will follow the strategy in the program below:

```
>let RURAL=Q041
>if Q041>1 then let RURAL=0
>if Q041=4 then let RURAL=.
```

5. Complete this first transformation by entering the Run command.

D. Categorize the per-pupil monetary, in-kind, and labour contributions of the community to school.

The next focus is contributions to school. The researchers who analyzed this data decided that given the disparity in size of schools in the sample, comparison of absolute contributions would be less useful to policy makers than comparison of contributions per-pupil. We will first create these per-pupil contribution rates, then categorize them.

1. Call up the data file you just saved (EXAMPL4A.SYS) and the additional variables E86(35) and E86(36) from the CENDATA.SYS file. These two additional variables record the total number of male pupils and female pupils (respectively) enrolled in the school. In the formulas below, we will add them to obtain the total number of pupils in the school. In this example, we will use the FORTRAN programming capabilities of SYSTAT to create new variables.

With the FORTRAN programming options, a few things are done differently:

1. Lines entered must be numbered.

2. Multiples of 10 can be used to leave space for adding in additional steps later.

3. Instead of introducing new variables just by entering them in an equation, SYSTAT must be warned they are coming by using a dimension (DIM) statement which opens space in the files memory for them. Since we will be creating new variables for per-pupil contributions for each of our three variables, we will dimension space for them.

4. Begin your program by entering the following:

```
>10 dim pps463(3)
```

This tells SYSTAT that you will need 3 variables:

pps463(1), pps463(2), and pps463(3).

The name pps463 is used as an aid to memory to indicate that these variables will be per-pupil analogs to the raw contribution data. We can now create those variables by entering the following lines to the program:

```
>20 let pps463(1)=s463a/(e86(35)+e86(36))
>30 let pps463(2)=s463b/(e86(35)+e86(36))
>40 let pps463(3)=s463c/(e86(35)+e86(36))
```

In a real analysis, we would specify a file for saving our new data file.

3. Run the program as it is, and repeat the steps of examining the distributions of the data. Since we are retracing the steps of experienced researchers who took pains to carefully plan their analysis, we will procede with the next step of categorizing. You can examine the distributions of the per-pupil contributions for additional practice in using SYSTAT on your own if you desire.

For the purposes of their subsequent analysis, the researchers decided to categorize the contributions as follows:

0 = 0 contribution or missing data
1 = 1 - 9 per pupil
2 = 10 - 19 per pupil
3 = 20 or more per pupil

We can create grouped versions of the variable by entering the following lines.

```
>50 dim monpp
>60 if pps463(1)=0 or pps463(1)=. then let monpp=0
>70 if pps463(1)>0 and pps463(1)<10 then let monpp=1
>80 if pps463(1)>10 and pps463(1)<20 then let monpp=2
>90 if pps463(1)>20 then let monpp=3
>100 dim labpp
>110 if pps463(2)=0 or pps463(2)=. then let labpp=0
>120 if pps463(2)>0 and pps463(2)<10 then let labpp=1
>130 if pps463(2)>10 and pps463(2)<20 then let labpp=2
>140 if pps463(2)>20 then let labpp=3
>150 dim inkpp
>160 if pps463(3)=0 or pps463(3)=. then let inkpp=0
>170 if pps463(3)>0 and pps463(3)<10 then let inkpp=1
>180 if pps463(3)>10 and pps463(3)<20 then let inkpp=2
>190 if pps463(3)>20 then let inkpp=3
```

You have now dimensioned 3 new variables:

MONPP, monetary contributions per pupil
LABPP labour contributions per pupil
INKPP, in-kind material contributions per pupil.

4. To conclude your work, enter a save command (the name EXAMPL4C.SYS will keep our system of organization intact), and run the program.

Data Sources for the

BRIDGES Sri Lanka Data Set

The data base for these examples and assignments consists of data for 273 schools drawn from 3 principal data sources. Within the sample of schools, 252 are under the control of the Ministry of Education; the remaining 21 schools are private. All data is available for the MOE controlled schools; some data may be missing for the independently run private schools.

The first data source is the Annual School Census. Data is collected annually from all schools as part of the normal operations of the Ministry or Education. The data included in the BRIDGES data files includes enrollment, teacher information, and financial information. Data for the years 1984, 1985, and 1986 is available; most of the data in the example files is from 1986. Data from the School Census is contained in the file CENDATA.

The remaining data comes from two surveys administered as part of the BRIDGES studies on the results of MOE school management reforms. These instruments, administered to the principal's of the sample schools, gathered information regarding aspects of management practice and management reform implementation. They also collected a great deal of data about the principal's themselves, their characteristics, their training, their careers as educators and principals. The two instruments, the Principals Survey and Principals Questionnaire, are the data sources for the files PRSDATA and PRQDATA respectively.

Attached to this cover sheet are excerpts of the original survey instruments and of the codebooks constructed during the coding of the data for research.

CODEBOOK FOR
SRI LANKA PROJECT
PRINCIPAL'S SURVEY AND
PRINCIPAL'S QUESTIONNAIRE DATA FILE

1.0 COMMUNITY CHARACTERISTICS

1.1 RURAL/URBAN

Rural or Urban location

1 = Rural
2 = Suburban
3 = Urban

Variable: Q041

Source(s): Principal's Questionnaire item Q041.

1.2 PARENTS' OCCUPATIONS

Occupational distribution of parents.

1 = Mostly professional and white collar
2 = Mostly paid shop employees and workers
3 = Mostly rural farm families
4 = Mostly plantation workers
5 = Mostly settlers in rural colonies
6 = Mostly fishermen
7 = Combination of the above

Variable: Q042

Source(s): Principal's Questionnaire item Q042.

1.3 FAMILY ECONOMIC CONDITIONS

Economic level of parents.

1 = very good
2 = good
3 = average
4 = somewhat poor
5 = poor

Variable: Q043

Source(s): Principal's Questionnaire item Q043.

1.4 PARENTS' EDUCATIONAL LEVEL

Educational level of parents

1 = not completed the primary cycle
2 = completed primary cycle
3 = studied up to GCE (O level)
4 = studied up to GCE (A level)
5 = received a higher education

Variable: Q044

Source(s):Principal's Questionnaire

1.5 PROXIMITY OF HOME TO SCHOOL

Proximity of students' homes to the school. Question asks whether the majority of students' parents live within a radius of three Km from the school.

1 = yes
2 = no

Variable: Q045

Source(s):Principal's Questionnaire

1.6 SOCIO-ECONOMIC STATUS

Principal's estimation of the socio-economic status of the families whose children attend school derived from the items Q043 and Q044 noted above.

Variable: SES

Source: Principal's Questionnaire items Q043 and Q044. Q043 variables recoded so that the <u>higher</u> numbers indicate higher status. Both variables ranged 1 to 5. Sum of recoded variables; range of 2 to 10.

==========

2.0 SCHOOL SYSTEM ORGANIZATIION

2.1 CLUSTER AGE

This variable indicates the age of the the cluster. Coded as missing if school is not in a cluster, otherwise the date of cluster formation (all 1981 or later) for those schools in clusters.

Variable: S07

Source(s): Principal's Survey item S07; the question asks for the date of cluster formation if the school is in a cluster.

2.2 STATUS IN CLUSTER

This variable indicates the cluster status of the school.

0 = school not in a cluster
1 = Lesser school in a cluster
2 = Stronger school in a cluster
3 = Core school in a cluster

Variable: CLSTATUS

Source(s): Principal's Survey items S07 and S06, and item C(8) from school census data in file CEN86F1.SYS.

2.3 CLUSTER AND BACKING

This variable indicates the schools rural or urban location, cluster status, and administrative support.

1 = Rural school not in a cluster
2 = Urban school not in a cluster
3 = Rural school in one of the original 'classic' clusters created before the 1984 reforms
4 = Urban school in one of the original 'classic' clusters created before the 1984 reforms
5 = Rural school in a modified cluster
6 = Urban school in a modified cluster

Variable: NCLUSTER

Source(s): Principal's Questionnaire item Q041 which inquires into the school's location (rural or urban); Principal's Survey item S07 which elicits the age of the cluster if the school is located in one.

2.4 CLUSTER MEMBERSHIP

This variable indicates whether or not a school is in a cluster.

0 = No, school is not in a cluster
1 = Yes, school is in a cluster

Variable: CLUYES

Source: Variable NCLUSTER described above.

2.5 DIVISION BACKING

This variable indicates the presence of a division office backing the cluster organization.

0 = No, no division office
1 = Yes, division office exists

Variable: DIVYES

Source: Variable NCLUSTER described above.

==========

3.0 COMMUNICATION AND COOPERATION

3.1 UPWARD COMMUNICATION

This variable reflects the frequency of communication from the school upwards to the division office and the Ministry of Education. Possible range is 0 to 20, highest observation at 18. All components correlate with COMUP (.266 to .620)

Variable: COMUP

Source: Principal's Survey and Principal's Questionnaire items S58, Q0112a, Q0112b, Q0112d. The values of S58 were reversed so that high communication = 5. COMUP is the sum of S58 (recoded), Q0112a, Q0112b, Q0112d.

3.2 DOWNWARD COMMUNICATION

This variable reflects the frequency of communication to the school from higher levels in the administrative system. Range from 0 to 20.

Variable: COMDOWN

Source: Principal's Survey and Principal's Questionnaire items S60, Q0111a, Q0111b, Q0111d. COMDOWN is the sum of these variables.

3.3 LATERAL COMMUNICATION

This variable reflects the frequency of communication between the school's principal and the principals of nearby schools. Possible range is from 0 to 12.

Variable: LATERAL

Source: Principal's Survey and Principal's Questionnaire items S564, Q0111c, Q0111d. S564 is reversed so the maximum value is 4. LATERAL is the sum of S564 (recoded), Q0111c, and Q0111d.

3.4 IN-SERVICE TRAINING FOR TEACHERS

Indicates whether or not the district helps teachers in developing skills.

1 = teachers receive help,
0 = teachers receive no help.

Variable: INSERVE

Source: Principal's Survey item S61b

3.5 TEACHER PARTICIPATION IN INTER-SCHOOL SEMINARS

Variable indicates the number of times in a year that teachers from a school attend seminars on teaching together with teachers from other schools

Variable: S211

Source: Principal's Survey item S211.

3.6 RESOURCE AUGMENTATION

This variable reflects the extent to which cluster formation has led to augmenting of school resources. Possible range from 0 to 13; all components are correlated with RESOURCE.

Variable: RESOURCE

Source: Principal's Survey items S49a-f, S482, S561, and S631. Each resource mentioned in S49 is given a value of 1, values from S482 and S631 are subracted from 2 and the result multiplied by 2 (e.g. 2*(2-S482)). The value of S561 is reversed (4-S561).

==========

4.0 PRINCIPAL'S CHARACTERISTICS

4.1 PRINCIPAL'S STATUS

Principal's Status;

1 = permanent
2 = acting
3 = performing

Variable: Q021

Source(s): Principal's Questionnaire item Q021.

4.2 PRINCIPAL'S LENGTH OF SERVICE

Principal's years of service at this school.

Variable: Q022

Source: Principal's Questionnaire item Q022

4.3 PRINCIPAL'S SERVICE STATUS

This item gives the principal's service status.

1 = SLEAS I
2 = SLEAS II
3 = SLEAS III
4 = Principal's Service I
5 = Principal's Service II
6 = Principal's Service III
7 = Other posts

Variable: Q028

Source: Principal's Questionnaire item Q028

4.4 PRINCIPAL'S QUALIFICATIONS

Shows level of principal's qualifications;

1 = Ph.D./M.A./M.Sc.
2 = B.Ed./B.A./B.Sc.
3 = GCE AL/GAQ/GSC
4 = GCE OL
5 = Other

Variable: Q031

Source: Principal's Questionnaire item Q031.

4.5 PRINCIPAL'S MANAGEMENT TRAINING

Item asks whether the principal has received special training in school management. 1 = yes, 2 = no.

Variable: Q033

Source: Principal's Questionnaire item Q033

4.6 PRINCIPAL'S TEACHING EXPERIENCE

Variable indicates the number of years a principal worked as a teacher without doing any administrative work.

Variable: Q026

Source: Principal's Questionnaire item Q026.

4.7 PRINCIPAL'S AGE

Variable: Q0141

Source: Principal's Questionnaire item Q0141

4.8 PRINCIPAL'S GENDER

Principal's gender. 1 = female, 2 = male.

Variable: Q0142

Source: Principal's Questionnaire item Q0142

4.9 PROXIMITY OF PRINCIPAL'S HOME

Distance of principal's home from school.

1 = 0 to 3 kilometers
2 = 4 to 8 kilometers
3 = more than 8 kilometers

Variable: Q0145

Source: Principal's Questionnaire item Q0145

4.10 PROXIMITY OF PRINCIPAL'S BIRTHPLACE

Proximity of principal's place of birth to the school.

1 = very close
2 = within the district
3 = in a neighbouring district
4 = further away

Variable: Q0147

Source: Principal's Questionaire item Q0147

4.11 PRINCIPAL'S TENURE AT THE SCHOOL

Principal's tenure at the school in years.

Variable: Q024

Source: Principal's Questionnaire item Q024.

5.0 PRINCIPAL'S MANAGEMENT PRACTICES

5.1 SHARED DECISION-MAKING

This variable measures the extent to which the principal includes school staff members in decision making.

Variable: DC4

Source(s): Principal's Survey items:

S50 Do you have a general staff meeting? (yes/no)
S51 How frequently does it meet? (1 to 7; 1=once or twice per week, 6=once per year)
S52 In this meeting, is there a free exchange of opinions? (yes/no)

DC4 consists of the sum of the recoded versions of the above variables where:

S50 recoded as DC1 where 1=Yes, 0=No, .5= missing
S51 recoded as DC2 where 1= 1to 4, 0= 7(other) .5= missing.
S52 recoded as DC3 where 1=Yes, 0=No, .5= missing.

5.2 INSTRUCTIONAL MANAGEMENT

IM4 indicates the extent to which the school principal observes and contributes towards teachers professional development through consultation, teaching demonstrations, and other means.

Variable: IM4

Sources: Principal's Survey items:

S15 How often do you visit the classess of your teachers?

1= several times per day
2= about once each day
3= two to three times each week
4= every fifteen days
5= once per month
6= once a year.

S18 How often do you suggest ways of improving student achievement to your teachers?

1= very often
2= often
3= infrequently
4= never

Q065A How much of the principal's time (as percent) is spent on supervising teachers?

Q067A How much time (in percent) does the principal spend talking to teachers?

IM4 was computed as the sum of the recoded the variables above; recoding as follows:

S15 recoded as IM1 where 1=1,2,or3; 0=4,5,or6; .5= missing.
S18 recoded as IM2 where 1=1,2,3; 0=4,5,or6; .5 missing.
Q065A recoded as Q065x where 0=1to9%; 1=10to80%; .5= missing. (10% is median value)
Q067A recoded as Q067x where 0=1to4%; 1=5to80%; .5=missing. (5% is median value)

5.3 PLANNING

This variable indicates the extent to which the principal develops a formal plan for the school.

Variable: PL4

Source: Principal's Survey items

S90 Has your school drafted a formal implementation plan? (yes/no)
S91 Has your school drafted a formal school development plan? (yes/no)
S93 Have you listed your expectations of the annual implementation plan? (yes/no)

PL4 is the sum of these variables.

5.4 STUDENT WELFARE ACTIONS

This variable indicates the extent to which the principal involves himself or herself in student welfare activities.

Variable: WF6

Sources: Principal's Survey items

S40.1 Do you get in touch with parents if students fail to come to school?
S41.2 Do you try to make special arrangements for students who miss classes for good reasons?
S45 Does your school have designed programmes for the development of sports activities or clubs?
S452c Is the principal himself involved in the

organizing of these activities?

S453e Does the principal himself participate in these activities?

WF6 is the sum of these variables.

5.5 CURRICULAR DEVELOPMENT

This variable indicates the extent to which the principal is active in curriculum development.

Variable: PC7

Source: Principal's Survey items S13.1, S13.5, S13.8, S13.9a, S13.10, S13.21. These items asked the follwing questions:

S13.1 Is there a standing curriculum committee?
S13.5 Do you think it is desirable for you to ammend or supplement the standard curriculum?
S13.8 To what extent is the in-school development of materials encouraged? (encouraged, or not)
S13.9a How often do the teachers of the same grade level meet to discuss their programme in implementing the curriculum? (often, or not)
S13.10 Do you meet with them?
S13.21 If there is a standing curriculum development committee, does the principal participate?

PC7 is the sum of these variables.

==========

6.0 INNOVATIONS

6.1 STUDENT CONDITIONS REFORMS

Number of 9 possible improvements in student conditions reported as having been implemented.

Variable:STUR

Source(s):Principal's Questionnaire items Q811a to Q819a.

6.2 STUDENT CONDITIONS REFORMS (proportional)

Proportion of 9 possible improvements in student conditions

reported as having been implemented.

Variable:STURP

Source(s):Principal's Questionnaire items Q811a to Q819a.

6.3 TEACHER DEVELOPMENT ACTIONS

Number of 8 possible actions to encourage teacher development reported as having been implemented.

Variable: TEAR

Source(s):Principal's Questionnaire items Q821a to Q828a.

6.4 TEACHER DEVELOPMENT ACTIONS (proportional)

Proportion of 8 possible actions to encourage teacher development reported as having been implemented.

Variable:TEARP

Source(s):Principal's Questionnaire items Q821a to Q828a.

6.5 TEACHING MATERIALS

Number of 2 possible provisions regarding teaching materials reported as having been implemented.

Variable:MATR

Source(s):Principal's Questionnaire items Q831a to Q832a.

6.6 TEACHING MATERIALS (proportional)

Proportion of 2 possible provisions regarding teaching materials reported as having been implemented.

Variable:MATRP

Source(s):Principal's Questionnaire items Q831a to Q832a.

6.7 BUILDING IMPROVEMENTS

Number of 3 possible improvements in buildings or plans for improvements reported as having been implemented.

Variable: BLDR

Source(s):Principal's Questionnaire items Q841a to Q843a.

6.8 BUILDING IMPROVEMENTS (proportional)

Proportion of 3 possible improvements in buildings or plans for improvements reported as having been implemented.

Variable: BLDRP

Source(s):Principal's Questionnaire items Q841a to Q843a.

6.9 COCURRICULAR IMPROVEMENTS

Number of 4 possible improvements in cocurricular activities facilities reported as having been implemented.

Variable: COCR

Source(s):Principal's Questionnaire items Q851a to Q854a.

6.10 COCURRICULAR IMPROVEMENTS (proportional)

Proportion of 4 possible improvements in cocurricular activities facilities reported as having been implemented.

Variable: COCRP

Source(s):Principal's Questionnaire items Q851a to Q854a.

6.11 STUDENT GUIDANCE

Number of 2 possible improvements in student guidance practices reported as having been implemented.

Variable: GUIR

Source(s):Principal's Questionnaire items Q861a to Q862a.

Proportion of 2 possible improvements in student guidance practices reported as having been implemented.

Variable: GUIRP

Source(s):Principal's Questionnaire items Q861a to Q862a.

6.13 SCHOOL COMMUNITY RELATIONS

Number of 2 possible means to improving school community relations reported as having been implemented.

Variable: SCRR

Source(s):Principal's Questionnaire items Q871a to Q872a.

6.14 SCHOOL COMMUNITY RELATIONS (proportional)

Proportion of 2 possible means to improving school community relations reported as having been implemented.

Variable: SCRRP

Source(s):Principal's Questionnaire items Q871a to Q872a.

==========

7.0 SCHOOL COMMUNITY RELATIONS

7.1 TOTAL PER CHILD COMMUNITY TO SCHOOL CONTRIBUTIONS

Variable: PK463

Source: Principal's Survey item S463a,b,c, and 1986 School Census Data enrollment figures E86(35) and E86(36).
PK463=(S463a+S463b+S463C)/(E85(35(+E86(36))

7.2 COMMUNITY TO SCHOOL MONETARY CONTRIBUTIONS PER CHILD

Variable: PKS463A

Sources: See above; S463a/enrollment.

7.3 COMMUNITY TO SCHOOL LABOR CONTRIBUTIONS PER CHILD

Variable: PKS463B

Sources: See above; S463b/enrollment.

7.4 COMMUNITY TO SCHOOL IN-KIND CONTRIBUTIONS PER CHILD

Variable: PKS463C

Sources: See above; S463c/enrollment.

7.5 TOTAL SCHOOL TO COMMUNITY CONTRIBUTIONS PER CHILD

Variable: PK473

Sources: Principal's Survey items S473a,b,c, and 1986 School Census Data items E86(35) and E86(36). PK473=(S473a+S473b+s473c)/(E86(35)+E86(36)).

7.6 TOTAL SCHOOL TO COMMUNITY MONETARY CONTRIBUTIONS PER CHILD

Variable: PKS473A

Sources: See above; S473a/enrollment.

7.8 TOTAL SCHOOL TO COMMUNITY LABOR CONTRIBUTIONS PER CHILD

Variable: PKS473B

Sources: See above; S473b/enrollment

7.9 TOTAL SCHOOL TO COMMUNITY IN-KIND CONTRIBUTIONS PER CHILD

Variable: PKS473C

Sources: See above; S473c/enrollment.

7.10 GROUPED TOTAL COMMUNITY TO SCHOOL CONTRIBUTIONS PER CHILD

Variable: GPK463

Source: Variable PK463 above;

0 = 0 or missing
1 = 1-9 Rs
2 = 10-19 Rs
3 = 20 or more Rs

7.11 GROUPED COMMUNITY TO SCHOOL MONETARY CONTRIBUTIONS PER CHILD

Variable: GPKS463A

Source: Variable PKS463A above;

0 = 0 or missing
1 = 1-9 Rs
2 = 10-19 Rs
3 = 20 or more Rs

7.12 GROUPED COMMUNITY TO SCHOOL LABOR CONTRIBUTIONS PER CHILD

Variable: GPKS463B

Source: Variable PKS463B above;

0 = 0 or missing
1 = 1-9 Rs
2 = 10-19 Rs
3 = 20 or more Rs

7.13 GROUPED COMMUNITY TO SCHOOL IN-KIND CONTRIBUTIONS PER CHILD

Variable: GPKS463C

Source: Variable PKS463C above;

0 = 0 or missing
1 = 1-9 Rs
2 = 10-19 Rs
3 = 20 or more Rs

7.14 GROUPED TOTAL SCHOOL TO COMMUNITY CONTRIBUTIONS PER CHILD

Variable: GPK473

Source: Variable PK473 above;

0 = 0 or missing
1 = 1-9 Rs
2 = 10-19 Rs
3 = 20 or more Rs

7.15 GROUPED SCHOOL TO COMMUNITY MONETARY CONTRIBUTIONS PER CHILD

Variable: GPKS473A

Source: Variable PKS473A above;

0 = 0 or missing
1 = 1-9 Rs
2 = 10-19 Rs
3 = 20 or more Rs

7.16 GROUPED SCHOOL TO COMMUNITY LABOR CONTRIBUTIONS PER CHILD

Variable: GPKS473B

Source: Variable PKS473B above;

0 = 0 or missing
1 = 1-9 Rs
2 = 10-19 Rs
3 = 20 or more Rs

7.17 GROUPED SCHOOL TO COMMUNITY IN-KIND CONTRIBUTIONS PER CHILD

Variable: GPKS473C

Source: Variable PKS473C above;

0 = 0 or missing
1 = 1-9 Rs
2 = 10-19 Rs
3 = 20 or more Rs

7.18 ATTENDANCE AT SCHOOL DEVELOPMENT SOCIETY MEETINGS

Question asks principal to indicate what percentage of School Development Society members usually attend meetings.

Variable: Q103

Source: Principal's Questionnaire item Q103

1 = 75-100%
2 = 50-75%
3 = 25-50%
4 = Less than 25%

7.19 SCHOOL DEVELOPMENT SOCIETY ACTIVITIES TO BENEFIT COMMUNITY

Variables: Q105A, Q105B, Q105C, Q105D, Q105E

Source: Principal's Questionnaire

7.20 SCHOOL DEVELOPMENT SOCIETY CULTURAL & DEVELOPMENT ACTIVITIES

Variables: Q106A, Q106B, Q106C, Q106D, Q106E, Q106F

Source: Principal's Questionnaire

==========

8.0 SCHOOL OUTPUTS

8.1 STUDENT ACHIEVEMENT (principal's estimation)

Principal's estimation of student achievement relative to national competition. 1 is high.

Variable: Q055

Source(s):Principal's Questionnaire item Q055, which asks;

With regard to student academic achievement how would you rate the school among all other schools in the nation?

1 = among the best
2 = better than average
3 = about average
4 = below average

5 = inferior

8.2 STUDENT ACHIVEMENT (principal's estimation)

Principal's estimation of student achievement relative to resourses. 1 is high

Variable: Q056

Source(s):Principal's Questionnaire item Q056, which asks;

With regard to student academic achievement how good a school do you think this school can be given the resources you have now?

1 = even though it is now among the best, it could be even better
2 = among the best
3 = better than average
4 = about average
5 = below average
6 = inferior

8.3 COCURRICULAR ACTIVITIES QUALITY

Principal's estimation of school's co-curricular activities.

Variable: Q057

Source(s):Principal's Questionnaire item Q057. which asks;

With regard to success in co-curricular activities how would you rate this school on a national basis?

1 = among the best
2 = better than average
3 = about average
4 = below average
5 = inferior

8.4 SCHOOL'S COMMUNITY SERVICE ORIENTATION (principal's estimation)

Principal's estimation of school community relations in relation to other schools inthe nation.

Variable: Q058

Source(s):Principal's Questionnaire item Q058, which asks;

With regard to helping the community, how would you rate this school on a national basis?

1 = among the best
2 = better than average
3 = about average
4 = below average
5 = inferior

8.5 SCHOOL EFFECTIVENESS (principal's estimation)

Principal's estimation of school's overall effectiveness.

Variable: S642

Source(s):Principal's Survey item S642, which asks;

According to your definition, how would you categorize your school?

1 = very efective
2 = effective
3 = average
4 = ineffective
5 = very ineffective

8.6 DAILY SCHOOL ATTENDANCE (principal's estimation)

Principal's appraisal of daily attendance.

Variable: Q072a

Source(s): Principal's Questionnaire item Q072a which asks the principal to estimate the average daily attendance.

1 = over 98%
2 = 90 to 97%
3 = 80 to 89%
4 = 70 to 79%
5 = 60 to 69%
6 = below 60%

8.7 DISCIPLINE

Number of kids disciplined by principal in a typical week.

Variable: Q072b

Source(s): Principal's Questionnaire item Q072b which asks the principal the number ofstudents brought to his/her office for dicipline in a typical week.

8.8 PROPORTION OF STUDENTS DISCIPLINED

Number of students reported disciplined in a week in the question immediately preceding expressed as a proportion of the school's enrollment.

Variable: PROPPUN

Sources: Principal's Questionnaire item Q072b which asks the principal the number ofstudents brought to his/her office for dicipline in a typical week, and the school's enrollment.

8.9 STUDENTS' ATTITUDES TOWARD SCHOOL

Principal's estimation of the percentage of kids who like school.

Variable: Q072c

Source(s):Principal's Questionnaire item Q072c which asks the principal to estimate the percentage of primary students who like school.

8.10 STUDENTS' ATTITUDES TOWARD SCHOOL

Principal's estimation of the percentage of kids who dislike school.

Variable: Q072d

Source(s): Principal's Questionnaire item Q072d which asks the principalto estimate the percentage of primary students who dislike school.

8.11 FIFTH GRADERS' NEXT STEP

Principal's estimation of proportion of school's fifth graders who went to National School. (Numerous errors in entry)

Variable: I1

Source(s):Principal's Questionnaire Item Q072I1

Principal's estimation of the proportion of school's fifth graders who went to a popular school. (Numerous errors in entry)

Variable: I2

Source(s):Principal's Questionnaire Item Q072I2

Principal's estimation of proportion of school's fifth graders who went to another school. (Numerous errors in entry)

Variable: I3

Source(s):Principal's Questionnaire Item Q072I3

Principal's estimation of proportion of school's fifth graders who went to this school. (Numerous errors in entry)

Variable: I4

Source(s):Principal's Questionnaire Item Q072I4

Principal's estimation of proportion of school's fifth graders who repeat. (Numerous errors in entry)

Variable: I5

Source(s):Principal's Questionnaire Item Q072I5

Principal's estimation of proportion of school's fifth graders drop out. (Numerous errors in entry)

Variable: I6

Source(s):Principal's Questionnaire Item Q072I6

Total proportion of school's fifth graders accounted for in the principal's estimation. (Numerous errors in entries)

Variable: I7

Source(s):Principal's Questionnaire Item Q072I1 to Q072I6.

Excerpts from
Project BRIDGES

PRINCIPAL'S SURVEY

Cover Questions

S06 Status in Cluster

S07 Date of incorporation in cluster. Month____ Year

Staff Meetings

5.0 Do you have general staff meetings?

1 = yes 2 = no

5.1 How frequently does it meet?

1 = one to two times a week
2 = once every fifteen days
3 = once every month
4 = once every three months
5 = once every six months
6 = once a year
7 = other

5.2 In this meeting, is there a free exchange of opinions?

1 = yes 2 = no

Planning

9.1 Has your school drafted a formal plan;

(i) Annual implementation plan for the school

1 = yes 2 = no

(ii) School development plan

1 = yes 2 = no

9.3 Have you listed your expectations of the annual implementation plan?

1 = yes 2 = no

Instructional Management

13. Now we would like to explore the extent to which the school develops an appropriate curriculum and the extent of your involvement in curricular development and in instructional organization. Could you describe the general approach in your school for developing the curriculum? Mark with an 'X' in the appropriate column.

13.1 Is there a standing curriculum development committee?

1. Yes, one for the whole school ____

2. Yes, different committees for each level; primary, junior secondary, senior secondary ____

3. No, curriculum management is handled by the principal and vice-principal. ____

13.2 If there are committees, who participates?

1. Principal ____

2. Teachers ____

3. Subject Co-ordinator ____

4. Community/SDS representatives ____

5. Other (Who?) ____

13.5 Do you think it is desirable for you to ammend or supplement the standard curriculum?

1 = yes 2 = no

13.8 To what extent is the in-school development of materials encouraged?

Very encouraged? 1 = yes 2 = no

Not encouraged at all? 1 = yes 2 = no

13.9 How often do the teachers of the same grade level meet to discuss their programme in implementing the curriculum?

Very often? 1 = yes 2 = no

13.10 Do you meet with them?

1 = yes 2 = no

15. How often do you visit the classes of your teachers?

a. Several times per day ____

b. About one time per day ____

c. Two to three times per week ____

d. Every fifteen days ____

e. Once a month ____

f. Once a year ____

18. How often do you suggest ways of improving student achievement to your teachers?

a. Very often ____

b. Often ____

c. Sometimes ____

d. Seldom ____

e. Never ____

20. How many times over the year will there be seminars or other sessions where groups of teachers discuss teaching practices?

____ times per year

21. Do the teachers of your school join with teachers from nearby schools in seminars?

1 = yes 2 = no

21.1 How often a year? ____ times per year

40.1 Do you get in touch with parent if students fail to come to school?

1 = yes 2 = no

41.2 Do you try to make special arrangements for students who miss classes for good reasons?

1 = yes 2 = no

45. Does your school have designed programmes for the development of sport activities of clubs?

1 = yes 2 = no

Who is involved in organizing these activities?

a. Physical Education teachers ____
b. Parents ____
c. Principal himself ____
d. Student Committees ____
e. other ____

Who participates in these activities?

a. All school staff ____
b. Community and parents ____
c. Students ____
d. Teachers ____
e. Principal himself ____
f. other ____

Community Involvement

46.3 What is the approximate value of the support rendered to the school by the community in the year 1986.

a. Monetary donations Rs. ____

b. Value of voluntary labour Rs. ____

c. Value of material donations Rs. ____

47.3 What is the approximate value of the support rendered to the community by the school in the year 1986.

a. Monetary donations Rs. ____

b. Value of voluntary labour Rs. ____

c. Value of material donations Rs. ____

School Reforms

48.1 How long has your school been involved in the cluster/zone/district reform agreement?

____ number of months

48.2 Has this new arrangement increased access to vital instructional resources?

1 = yes 2 = no

49. Has the lead cluster school provided more

a.	Teachers	1
b.	Visual aids	2
c.	Equipment	3
d.	Materials	4
e.	Stationery	5
f.	Textbooks	6

56. To what extent do you agree with the following statements regarding the differences you have observed as a result of this new arrangement? (State your level of agreement.)

(i) I believe that schools in the cluster/zone are in an advantageous position than before in sharing resources with other schools.

1 = Thoroughly agree
2 = Agree
3 = Disagree
4 = Thoroughly disagree

56.4 With the introduction of the cluster system, principals in nearby schools in the cluster like to work together to achieve common goals.

1 = Thoroughly agree
2 = Agree
3 = Disagree
4 = Thoroughly disagree

58. What is the frequency with which you visit the district office?

1 = One time per day
2 = Two or three times per week
3 = Every 15 days
4 = Once a month
5 = Two or three times per year

60. How many times does somebody from the district office visit your school?

1 = One time per day
2 = Two or three times per week
3 = Every 15 days
4 = Once a month
5 = Two or three times per year

61. Does the district office provide useful guidance for ;

a.	Principals	1 = yes	2 = no
b.	Teachers	1 = yes	2 = no
c.	Others	1 = yes	2 = no

63.1 Is there a real sharing of resources, i.e. educational materials, teachers, scholarships, etc., between your school and other schools in the cluster?

1 = yes 2 = no

64.2 According to your own definition of effectiveness, please note in a scale of five points, 1 being very effective, 5 being very ineffective, how you would categorise your school.

1 = very effective
2 = effective
3 = average
4 = ineffective
5 = very ineffective

Excerpts from
Project BRIDGES

PRINCIPAL'S QUESTIONNAIRE

2. Cover

2.1 What is your present status?

1 = Permanent Principal
2 = Acting Principal
3 = Performing Principal

2.2 How long have you been principal of this school?

_____ (number of years)

2.4 Altogether, how many years have you served as principal?

_____ (number of years)

2.6 How many years were you a principal without doing any administrative work?

_____ (number of years)

2.8 What is your present service status?

1 = SLEAS I
2 = SLEAS II
3 = SLEAS III
4 = Principal's Service I
5 = Principal's Service II
6 = Principal's Service III
7 = Other (please specify)

3. Education and Training

3.1 What is the highest qualification you have obtained?

1 = PhD/MA/MSc
2 = BEd/BA/BSc
3 = GCE AL/GAQ/GSQ
4 = GCE OL
5 = Other

3.3 Have you received any special training on school management?

Yes No

Yes No

4. School Environment

4.1 In what type of area is your school located?

1 = Rural
2 = Suburban
3 = Urban

4.2 What of the following best describes the type of families who send children to your school?

1 = Mostly professional and white collar
2 = Mostly paid shop employees and workers
3 = Mostly rural farm workers
4 = Mostly settlers in rural colonies
5 = Mostly fishermen
6 = A mix of some of the above

4.3 Would you say that the economic conditions of the majority of the parents of the students attending the school are

1 = very good
2 = Good
3 = Average
4 = Somewhat poor
5 = Poor

4.4 What do you think of the educational status of the majority of the parents of your students? Circle the highest level of the majority's attainment level.

1 = Not completed the primary cycle
2 = Completed the primary cycle
3 = Studied up to GCE (O level)
4 = Studied up to GCE (A level)
5 = Received a higher education

4.5 Thinking of where the parents of your students live, do a majority live within a radius of 3 Km. from the school?

Yes No

5.

5.5 With regard to student academic achievement, how would you rate the school among all other schools in the nation?

1 = Among the best
2 = Better than average

5.6 With regard to student achievement, how good a school do you think this school can be given the resources you now have?

1 = Even though it is among the best, it could be even better
2 = Among the best
3 = Better than average
4 = About average
5 = Below average
6 = Inferior

5.7 With regard to success in co-curricular activities, how would you rate this school on a national basis?

1 = Among the best
2 = Better than average
3 = About average
4 = Below average
5 = Inferior

5.8 With regard to helping the community, how would you rate this school on a national basis?

1 = Among the best
2 = Better than average
3 = About average
4 = Below average
5 = Inferior

7. Student Performance and Welfare

7.2 Do you have students in years 1-5?

Yes No

If no, please go to question 7.?.

If yes, please answer the following questions.

a. About what is the average daily attendance in your school?

1 = Over 98%
2 = 90% to 97%
3 = 80% to 89%
4 = 70% to 79%
5 = 60% to 69%
6 = Below 60%

b. In a typical week, approximately how many children from year 1 to 5 are brought to your office for disciplinary reasons?

(number of students)

b. In a typical week, approximately how many children from year 1 to 5 are brought to your office for disciplinary reasons?

_____ (number of students)

c. About what percent of the primary year children seem to like school?

_____ (percent estimated)

d. About what percent seem to dislike school?

_____ (percent estimated)

i. What happened to your fifth year class/classes last year? Please indicate the percent who did each of the following.

_____ Entered sixth year in a national school

_____ Entered sixth year in a popular school

_____ Entered sixth year in another school

_____ Stayed here for the sixth year

_____ Repeated the fifth year

_____ Discontinued their schooling

8. New Projects

8.1 Below is a list of projects suggested by the M.O.E. in a circular. If you attempted any of these projects, please list the year when you began, and indicate your judgement of success. (1 = very successful; 2 = somewhat successful; 3 = not successful.)

	Project	Year begun	Success
8.1.1	Participation		
8.1.2	Retention		
8.1.3	Drop-oputs		
8.1.4	Pupil absenteeism		
8.1.5	Late arrivals		
8.1.6	Student discipline		
8.1.7	Achievement levels		
8.1.8	Nutrition		
8.1.9	Medical inspection and follow-up		
8.2.1	Teacher requirements		
8.2.2	Teacher utilization		
8.2.3	Teacher retention		
8.2.4	Teacher absenteeism		
8.2.5	Teacher punctuality		
8.2.6	In-service training of teachers (internal)		
8.2.7	In-service training of teachers (external)		
8.2.8	Professional growth of teachers		
8.3.1	Revision and development of teaching materials		
8.3.2	Special projects		

8.4.1 Repairs to existing buildings and infrastructure facilities

8.4.2 Assessment of future needs for facilities

8.4.3 S.D.S and community support for plant improvements

8.5.1 Sports and athletics

8.5.2 Scouting, cadeting, guiding, etc. (contact with external agencies needed)

8.5.3 Religious, cultural, and aesthetic activities

8.5.4 Clubs, societies, and associations

8.6.1 Pupil personality development records

8.6.2 Career guidance

8.7.1 School Development Society, Old Students' Associations, etc.

8.7.2 Parent, school, and community relations

11. Communications

11.1 Since January 1st, how many times have you personally visited:

	5 or more	3 to 4	1 to 2	0
Ministry of Education officials	4	3	2	1
District Education officials	4	3	2	1
Principals of nearby schools	4	3	2	1
Divisional Education officials	4	3	2	1

11.2 Since January 1st, how many times have you been visited by:

	5 or more	3 to 4	1 to 2	0
Ministry of Education officials	4	3	2	1
District Education officials	4	3	2	1
Principals of nearby schools	4	3	2	1
Divisional Education officials	4	3	2	1

CODEBOOK FOR
SRI LANKA PROJECT
SCHOOL CENSUS DATA FILE

1.0 COMMUNITY CHARACTERISTICS

1.1 CENSUS DISTRICT

District of Sri Lanka where community is located.

1 = Colombo
2 = Colombo (Homogama)
11 = Tangalle
17 = Batticaloa
27 = Polonaruwa
28 = Bandarawella
31 = Kegalle

Variable: C(4)

Source: School Census Data in file CEN86F1.SYS

1.2 DISTRICT

Adjusted categorization of community's district location collapsing categories 1 and 2 above. (Only one school, geographically near to Colombo, was present in category two.)

1 = Colombo (including one school in Homogama)
11 = Tangalle
17 = Batticaloa
27 = Polonaruwa
28 = Bandarawella
31 = Kegalle

Variable: DISTRICT

Source: School Census Data in file CEN86F1.SYS

1.3 ETHNICITY

Community's ethnic make-up.

1 = Singhalese
2 = Tamil
3 = Muslim

Variable: c(6)

Source(s): School Census Data in file CEN86F1.SYS

2.0 SCHOOL FEATURES

2.1 SCHOOL NUMBER

Identifying number for schools in the data set.

Variable: SCNO

Source: School Census Data 1986

2.2 SCHOOL TYPE

Type of school:

1 = Type 1A-B (prepare students for A-level, all courses)
2 = Type 1C (prepare students for A-level, not science course)
3 = Type 2 (prepare students for O-level)
4 = Type 3 (provide basic education)

Variable: C(8)

Source: School census data in file CEN86F1.SYS

2.3 SCHOOL TYPE

Condensed categorization of school type:

0 = Type 2 or 3
1 = Type 1A, 1B, or 1C
2 = Private schools

Variable: TYPE

Source: School census data file CEN86F1.SYS, variable C(8)

2.4 TOTAL ENROLLMENT

Total number of students enrolled in the school, K through grade 13, in 1986. (Note: Not all schools enroll students in all

grades; this variable records the total number of students in whatever grades a school enrolls.)

Variable: STU

Source: School census data file CEN86F2.SYS variables C(35) and C(36) (total enrollments for boys and girls)

2.5 TOTAL STUDENTS GRADES 1 TO 6 IN 1986

Total number of students in grades 1 through 6 in schools in 1986.

Variable: TOTSTU

Source(s): School census data from 1986 in the data file CEN86F2.SYS, sum of variables E86(3 to 14).

2.6 STUDENTS GRADES 1 TO 6 IN 1985

Numbers of students enrolled in grades 1 to 6 in 1985.

Variables: E85T(1-6) = Number of students in each of grades 1 through 6 in 1986.
E85TOT = total number of students enrolled in grades 1 through 6 in a school in 1985.

Source: School census data in file CEN85F2.SYS; sum of male and female enrollments in grades 1 to 6 (variables E85(3 to 14)).

2.7 MALE STUDENTS (grades 1 to 6)

Total number of male students in grades 1 through 6 in schools in 1986.

Variable: MALESTU

Source(s): School census data from 1986 in the data file CEN86F2.SYS, sum of variables with odd numbered sub-scripts from E86(3 to 14) (E86(3), E86(5), E86(7), and so on.)

2.8 FEMALE STUDENTS (grades 1 to 6)

Total number of female students in grades 1 through 6 in schools in 1986.

Variable: FEMSTU

Source(s): School census data from 1986 in the data file CEN86F2.SYS, sum of variables with even numbered subscripts from E86(3 to 14) (E86(4), E86(6), E86(8), and so on.)

2.9 PROPPRTION OF FEMALE STUDENTS

The proportion of a school's 1986 enrollment which was female.

Variable: PROPFEM

Source: School enrollment data in file CEN86F2.SYS.

2.10 REPEATERS IN 1986

Number of repeaters in grades 1 through 6 in 1986.

Variables: R86T(1-6) number of repeaters in each of grades 1 through 6.
R86TOT= total number of repeaters grades 1 to 6.

Source: 1986 School Cenesus Data in file 'CEN86F3.SYS', variables F86(3) to F86(14). Sum of male and female repeaters in grades 1 through 6 in 1986

2.11 DROPOUTS IN/AFTER 1985

Number of students calculated as having dropped out of grades 1 through 5 during or after the 1985 school year.

Variables: D85T(1-5) = number of students who apparently dropped out during or after each of grades 1 through 5 in 1985.
D85TOT = total number calculated as having dropped out during or after the 1985 school year, grades 1 through 5.

Source: 1985 and 1986 School Census Data. Number of dropouts computed from the number of repeaters and enrollees for 1985 and 1986.

2.12 TOTAL TEACHERS (grades 1 to 6)

Total number of teachers of grades 1 through 6 in schools in 1986.

Variable: TOTTCH

Source(s): School census data from 1986 in the data file CEN86F7.SYS, sum of variables J(1) -- male teaachers -- and J(5)-- female teachers-- in grades 1 through 6 in 1986.

2.13 TEACHER ABSENTEEISM

Total number of days absent for teachers in the school

Variable: TOTDAYAB

Source: 1986 School Census files CEN86F5 and CEN86F6. The census reports the number of teachers' absences in categories. This variable was constructed by mulitplying the number of teachers in a category (e.g. 1 to 5 days absent) by the midpoint of that interval. Sum of these products constitutes the total.

2.14 TEACHER ABSENCE RATE

Average number of days absent per teacher

Variable: TEACHABR

Source: Variable TOTDAYAB described above and the total number of teachers reported in census data (file CEN86F7.SYS, variables J86(4) and J86(8). TEACHABR=TOTDAYAB/(J86(4)+J86(8)).

2.15 STUDENT TEACHER RATIO

Ratio of students to teachers in grades 1 through 6 in schools in 1986.

Variable: STUTCH

Source(s): The variables TOTSTU and TOTTCH described above; original figures from CEN86F2.SYS and CEN86F7.SYS, STUTCH is computed as TOTSTU/TOTTCH.

2.16 TOTAL SCHOOL EXPENDITURES IN 1986

Total school expenditures reported in the school census in 1986.

Variable: EXPEND

Source: Variables in the school census data from 1986: Sum of variables D86(14), D86(29), D86(30) (from the file CEN86F1.SYS), and variable K86(4) (from the file CEN86F7.SYS).

2.17 EXPENDITURES PER STUDENT FOR 1986

Expenditures per student as reported in the 1986 school census.

Variable: UNIT

Source: Variable EXPEND described above; variable STU86 described above. UNIT=EXPEND/STU

3.0 SCHOOL COMMUNITY RELATIONS

3.1 FACILITIES FEES 1986

Facilities fees collected by the school according to 1986 census data.

Variable: D86(6)

Sources: 1986 School Census Data

3.2 TEACHER SALARIES 1986

Teachers salaries as reported in the 1986 school census data.

Variable: K86(4)

Source: 1986 School Census Data in file CEN86.F7.SYS

3.3 TOTAL SCHOOL DEVELOPMENT SOCIETY FUNDS 1984

Sum of funds from Student Development Societies as reported in the 1984 school census data.

Variable: SDS84

Source: 1984 School Census Data items D84(29), D84(30), D84(31), D84(32); sum of these items with missing values set to 0.

3.4 TOTAL SCHOOL DEVELOPMENT SOCIETY FUNDS 1985

Sum of funds from School Development Societies as reported in the 1985 school census data.

Variable: SDS85

Source: 1985 School Census Data items D85(29), D85(30), D85(31), D85(32); sum of these items with missing values set to 0.

3.5 TOTAL SCHOOL DEVELOPMENT SOCIETY FUNDS 1986

Sum of funds from School Development Societies as reported in the 1986 school census data.

Variable: SDS86

Source: 1986 School Census Data items D86(29), D86(30), D86(31), D86(32).

3.6 TOTAL SCHOOL FUNDS 1986

Total funds available for the school in 1986

Variable: TOTMON

Source: 1986 School Census Data; sum of SDS86, D86(6), and K86(4) with all missing values set to 0.

3.7 SCHOOL DEVELOPMENT SOCIETY FUNDS PER CHILD

School Development Society funds per child enrolled in the school.

Variable: PKSDS

Source: 1986 School Census Data: SDS86/(E86(35)+E86(36))

3.8 FACILITIES FES PER CHILD

Facilities fees collected per child enrolled in the school in 1986.

Variable: PKFF

Source: 1986 School Census Data; D86(6)/(E86(35)+E86(36))

3.9 TEACHER SALARIES PER CHILD

Expenditures for teacher salaries in 1986 per child enrolled.

Variable: PKTSAL

Source: 1986 School Census Data; K86(4)/(E86(35)+E86(36))

3.10 TOTAL FUNDS PER CHILD

Total funds available to the school per child enrolled in the school in 1986.

Variable: PKTOTMON

Source: 1986 School Census Data; TOTMON/(E86(35)+E86(36))

3.11 PROPORTION OF INCOME FROM FACILITIES FEES 1986

Proportion of a school's income in 1986 that came from facilities fees.

Variable: PROPFF

Source: 1986 School Census Data; D86(6)/TOTMON

3.12 PROPORTION OF INCOME FROM SCHOOL DEVELOPMENT SOCIETIES 1986

Proportion of a school's income in 1986 which came from School Development Societies.

Variable: PROPSDS

Source: 1986 School Census Data; SDS86/TOTMON.

3.13 PROPORTION OF FUNDING EXPENDED FOR TEACHER SALARIES 1986

Proportion of funding expended on teacher salaries.

Variable: PROPTSAL

Source: 1986 School Census Data; K86(4)/TOTMON.

3.14 GROWTH IN SCHOOL DEVELOPMENT SOCIETY FUNDS: 1984 TO 1986

Variable: GROW8486

Source: 1984 and 1986 School Census Data; SDS86/SDS84 with missing values for SDS84 set to 0.

3.15 GROWTH IN SCHOOL DEVELOPMENT SOCIETY FUNDS: 1985 TO 1986

Variable: GROW8586

Source: 1985 and 1986 School Census Data; SDS86/SDS85 with missing values for SDS85 set to 0.

3.16 GROWTH IN SCHOOL DEVELOPMENT SOCIETY FUNDS: 1984 TO 1985

Variable: GROW8485

Source: 1984 and 1985 School Census Data; SDS85/SDS84 with missing values for SDS84 set to 0.

3.17 TOTAL PER CHILD COMMUNITY TO SCHOOL CONTRIBUTIONS

Variable: PK463

Source: Principal's Survey item S463a,b,c, and 1986 School Census Data enrollment figures E86(35) and E86(36).
PK463=(S463a+S463b+S463C)/(E85(35(+E86(36))

3.18 COMMUNITY TO SCHOOL MONETARY CONTRIBUTIONS PER CHILD

Variable: PKS463A

Sources: See above; S463a/enrollment.

3.19 COMMUNITY TO SCHOOL LABOR CONTRIBUTIONS PER CHILD

Variable: PKS463B

Sources: See above; S463b/enrollment.

3.20 COMMUNITY TO SCHOOL IN-KIND CONTRIBUTIONS PER CHILD

Variable: PKS463C

Sources: See above; S463c/enrollment.

4.0 SCHOOL OUTCOMES

4.1 REPETITION RATES FOR 1985

Repetition rates for the primary grades; proportion of students who repeated a grade in 1986 as a percentage of the 1985 enrollment.

Variables:

RR85(1-5) = Repeaters in grades 1 through 5 in 1986 as a proportion of those enrolled in 1985;
e.g. RR85(1)=R86T(1)/E85T(1)

RR85AVG = The average repetition rate across the first five grades (sum RR85(1-5)/5).

RR85ADJ= Adjusted repetition rate; those cases with repetition rates or dropout rates which constitute statistical 'outliers' (RR85AVG>30, DR85AVG>.17) set to missing.

Source: School Census Data for 1985 and 1986. See variables E85T, R86T, etc. above.

4.2 DROPOUT RATES FOR 1985

Dropout rates for the primary grades; proportion of students who dropped out during or after 1985 as a percentage of the 1985 enrollment.

Variables:

DR85(1-4) = Dropouts from grades 1 through 4 in 1986 as a proportion of those enrolled in 1985;
e.g. DR85(1)=D85T(1)/E85T(1)

DR85AVG = The average dropout rate across the first four grades (sum DR85(1-4)/4).

DR85ADJ = Adjusted dropout rates; cases where

repetition rate or dropout rate constituted statistical 'outlier' (RR85AVG>.3, DR85AVG>.17) set to missing.

Source: School Census Data for 1985 and 1986. See variables E85T, D85T, etc. above.

4.3 GROWTH IN ENROLLMENT: 1984 to 1986

Growth rate of schools from 1984 to 1986.

Variable: GR84to86

Source: School Census Data from 1984, 1985, 1986.
GR84TO86=(E86TOT/E84TOT)*100

1985-1986 SCHOOL CENSUS DATA

LOCATION AND BUDGET

SCNO School number.

C86(4) Census District.

C86(6) Type by ethnic group; 1=Singhalese, 2=Tamil, 3=Muuslim.

C86(8) School type; 1 = type IA,IB (A-level preparation, all subjects); 2 = type IC (A-level preparation except science); 3 = type II (O-level preparation); 4 = type IV (basic education).

D86(6) Total facilities fees collected.

D86(14) Total expenditures.

D86(29) Total School Development Society expenditures.

D86(30) Total Old Student Association expenditures

D86(31) School Development Society balance on December 31.

D86(32) Old Student Association balance on December 31.

ENROLLMENT 1985

E85(3) Number of boys enrolled in grade 1.

E85(4) Number of girls enrolled in grade 1.

E85(5) Number of boys enrolled in grade 2.

E85(6) Number of girls enrolled in grade 2.

E85(7) Number of boys enrolled in grade 3.

E85(8) Number of girls enrolled in grade 3.

E85(9) Number of boys enrolled in grade 4.

E85(10) Number of girls enrolled in grade 4.

E85(11) Number of boys enrolled in grade 5.

E85(12) Number of girls enrolled in grade 5.

E85(13) Number of boys enrolled in grade 6.

E85(14) Number of girls enrolled in grade 6.

E85(35) Total number of boys enrolled, all grades, all levels.

E85(36) Total number of girls enrolled, all grades, all levels.

ENROLLMENT 1986

E86(3) Number of boys enrolled in grade 1.

E86(4) Number of girls enrolled in grade 1.

E86(5) Number of boys enrolled in grade 2.

E86(6) Number of girls enrolled in grade 2.

E86(7) Number of boys enrolled in grade 3.

E86(8) Number of girls enrolled in grade 3.

E86(9) Number of boys enrolled in grade 4.

E86(10) Number of girls enrolled in grade 4.

E86(11) Number of boys enrolled in grade 5.

E86(12) Number of girls enrolled in grade 5.

E86(13) Number of boys enrolled in grade 6.

E86(14) Number of girls enrolled in grade 6.

E86(35) Total number of boys enrolled, all grades, all levels.

E86(36) Total number of girls enrolled, all grades, all levels.

REPEATERS 1986

F86(3) Number of boys repeating grade 1.

F86(4) Number of girls repeating grade 1.

F86(5) Number of boys repeating grade 2.

F86(6) Number of girls repeating grade 2.

F86(7) Number of boys repeating grade 3.

F86(8) Number of girls repeating grade 3.

F86(9) Number of boys repeating grade 4.

F86(10) Number of girls repeating grade 4.

F86(11) Number of boys repeating grade 5.

F86(12) Number of girls repeating grade 5.

F86(13) Number of boys repeating grade 6.

F86(14) Number of girls repeating grade 6.

F86(35) Total number of boys repeating, all grades, all levels.

F86(36) Total number of girls repeating, all grades, all levels.

TEACHERS

H86(6) Number of teachers with 0 days on leave.

H86(12) Number of teachers with 1-4 days on leave.

H86(18) Number of teachers with 5-8 days on leave.

H86(24) Number of teachers with 9-12 days on leave.

H86(30) Number of teachers with 13-16 days on leave.

H86(36) Number of teachers with 17-20 days on leave.

I86(6) Number of teachers with 21-24 days on leave.

I86(12) Number of teachers with 25-28 days on leave.

I86(18) Number of teachers with 29-32 days on leave.

I86(24) Number of teachers with 33-36 days on leave.

I86(30) Number of teachers with 37-40 days on leave.

I86(36) Number of teachers with 41 or more days on leave.

I86(42) Total number of teachers.

J86(1) Number of male teachers, grades 0-5.

J86(2) Number of male teachers, grades 6-10.

J86(3)	Number of male teachers, grades 11-12
J86(4)	Total number of male teachers.
J86(5)	Number of female teachers, grades 0-5.
J86(6)	Number of female teachers, grades 6-10.
J86(7)	Number of female teachers, grades 11-12.
J86(8)	Total number of female teachers.
K86(1)	Total salaries for teachers, grades 0-5.
K86(2)	Total salaries for teachers, grades 6-10.
K86(3)	Total salaries for teachers, grades 11-12.
K86(4)	Total salaries, all teachers, all levels.

ASSIGNMENT #6: IDENTIFYING AND CREATING VARIABLES

PURPOSE: Assignments 6-9 will cover the basic steps in conducting and reporting a policy study. An important part of the assignments will be learning to use and merge various kinds of data in a central database for the purpose of research.

Assignment #6 will focus on (1) retrieving files from R:Base and loading them into LOTUS and (2) merging LOTUS files to create new variables.

DATE DUE: Tuesday morning, August 16

DIRECTIONS: Your group will be assigned a policy issue to analyze using data in the Sri Lanka database. To begin your analysis of the issue, your group should complete the following steps:

1. Identify the variables (independent, dependent, conditional) in your study.

2. Draw a diagram which shows the relationships between the variables. (See "Policy Analysis", Chapter 2, page .)

3. Using the CodeBook, identify the R:Base files needed for your analysis of these variables. (You may also need to use the copies of the data collection instruments to identify which files you need.)

4. Retrieve the variables from R:Base and into LOTUS. Print out a listing of the variables.

5. Create at least one new variable for your analysis by merging LOTUS files. Make a print-out of your results.

6. Bring your group's diagram and print-outs to class on Tuesday morning.

REFERENCES: "Policy Analysis" Chapter 2.

ASSIGNMENT #7: UNIVARIATE ANALYSIS AND CLEANING DATA

PURPOSE: This assignment focuses on the need to carefully examine individual variables (univariate analysis) before bivariate or multivariate analysis is begun. Failure to do this can result in unreliable findings and a loss of time. Univariate analysis can also alert the researcher that the data need to be "cleaned" before further analysis is conducted (because of data that are missing, incorrectly coded, etc.)

In this assignment, your group will focus on univariate analysis, data cleaning, and then will begin to examine the bivariate relations in your policy study.

DATE DUE: Wednesday morning, August 17

DIRECTIONS: Working with the variables for the policy issues your group has been assigned, complete the following steps:

1. Conduct univariate analysis on the variables for your study. Print out any frequency distributions or other tables that indicate how you have examined the variables.

2. Examine your data to see where it needs to be cleaned. Go into the original files and clean them as necessary. Write a brief paragraph to describe what was wrong with the data and what you did to clean it.

3. Complete any analyses that you did yesterday with the newly cleaned data (creating variables, etc.)

4. Finally, generate and print out (a minimum of two) new tables displaying the bivariate relationships between your variables. On Wednesday morning, bring to class:

   a. the results of your univariate analysis,
   b. the paragraph describing the data cleaning, and
   c. the two (or more) new tables of your variables.

REFERENCES: "Policy Analysis", Chapter 2

ASSIGNMENT #8: USING A STATISTICAL PACKAGE

PURPOSE: Spreadsheet packages like LOTUS are some of the most widely used types of software in the world. While they can certainly be used for statistical analysis, there are some advantages in using software specifically designed for that purpose.

This assignment will give you the opportunity to use a sophisticated statistical package (in this case, SYSTAT) for the same kinds of tasks covered in Assignments 6 and 7. Since you already understand the exercise from a conceptual level, focus your attention on learning as much as possible about SYSTAT and noting how it makes these statistical tasks easier. Would the flexibility and ease of use make it worth the extra cost for your Ministry?

DATE DUE: Thursday morning, August 18

DIRECTIONS: Follow the same instructions as for Assignments 6 and 7. However, this time:

1. Retrieve the files for your group's policy study from R:Base and load them into SYSTAT.

2. Complete the rest of the analyses as you did in LOTUS, only now in SYSTAT. Bring print-outs of your results to class Thursday morning.

REFERENCES: "Policy Analysis", Chapter 2, SYSTAT Manual.

ASSIGNMENT #9: CONDUCTING A POLICY STUDY (FINAL PROJECT)

PURPOSE: Unless the results of a policy study are presented clearly, the results may have little impact. Researchers must be skilled not only in proposal design and statistical analysis, but in ways of communicating their findings to others effectively.

This project will simulate the final stages of a policy study. Although much of your analysis has already been completed, your group will now focus on the presentation of the results. You will present the results in written form and in a brief oral presentation.

DATE DUE: Friday afternoon, August 19

DIRECTIONS: For the written portion of this assignment, your group should complete a brief report that covers each of the following topics:

1. Title of study

2. Statement of the problem

3. Significance of the problem

4. Audience to whom results will be reported and implications this has for design of study

5. Hypotheses/Research questions (Complete the matrix found on page 23, Chapter 2, "Policy Analysis".)

6. Diagram showing Relationships between variables: (See "Policy Analysis", Chapter 2, pg.)

7. Results - Write a paragraph or two summarizing your major findings.

8. Attach any diagrams, graphs, tables, etc. to the written portion of your report.

(Continued on the next page.)

For your oral report, your group should be prepared to give a not more than 10 minute presentation of your results. It may be helpful to have handouts for the class of some of your major tables (or transparencies can be made in the Instructional Technology Center).

REFERENCES: "Policy Analysis", Chapter 3.

Section IV: System for Tracking Educational Progress STEP

A USER'S GUIDE TO THE SYSTEM FOR TRACKING EDUCATIONAL PROGRESS

(STEP)

• AUGUST 1988 •

Developed by:
Luis Crouch and Scott Moreland
Research Triangle Institute

with:
Ernesto Cuadra and Noel McGinn
Harvard University

BRIDGES Project

Sponsored by the Office of Education, Bureau for Science and Technology,
and
Office of Development Resources, Bureau for Latin America and the Caribbean
United States Agency for International Development

The Basic Research and Implementation in DevelopinG Education Systems (BRIDGES) Project is directed by Harvard Institute for International Development and the Harvard Institute for International Development and the Harvard Graduate School of Education under Cooperative Agreement No. DPE-5824-A-5076 with the Office of Education, USAID.

TABLE OF CONTENTS

CHAPTER 1. INTRODUCTION TO THE STEP SYSTEM 1

Introduction to the STEP System 1

STEP Modules 1

Using STEP on the Microcomputer 1
Hardware Environment 1
Software Environment 2

Organization of this Manual 2

CHAPTER 2. USING STEP 3

Interactive vs. Batch Modes 3

File Management Under STEP 4

STEP Modules: POP 5
Description of the POP Module. 5

Use of the POP module 6
Specifying rates and initial values 6
Specifying mortality 7
Specifying International and Internal Migration 8
Base Year Population 8
Sex Ratio 9
Index of Females 9

STEP Modules: ESTIMATE 9
Description of the ESTIMATE Module 9
Use of the ESTIMATE Module 11

STEP Modules: PROJRATE 13
Description of the PROJRATE Module 13
Use of the PROJRATE Module 15

STEP Modules: EDPROJ 16
Description of the EDPROJ Module 16
Use of the EDPROJ Module 17

STEP Modules: COST 18
Description of the COST module 18
Use of the COST Module 19

STEP Modules: COHORT 20
Description of the COHORT Module 20
Use of the COHORT Module 20

Running the STEP System in Batch Mode 22

CHAPTER 3. CONFIGURING STEP TO A NEW APPLICATION 25

Step 1: Setting the Current Maximum Levels of Dimensions 25
Dimensions Used by STEP Variables in the STEP Data Base..... 26
Dimensions Used by POP Variables in the POP Data Base..... 26

Step 2: Defining and Editing the Label Library 27

Step 3: Entering New Data for the New Application 27

Step 4: Defining New Tables and Graphs 27

Variables in the STEP Data Files 28

Variables in the POP Data Files 29

Appendix to Chapter 3 31
Label Format Definitions 31

CHAPTER 4. TECHNICAL DESCRIPTION OF THE MODEL METHODOLOGY 39

The Grade-transition Enrollment Projection Method 39
Projections with Fixed Parameters 40
Projections with Variable Parameters 40
Linear Trend 40
Logistic Curve Fitting 40
Target Setting Method 41

Efficiency Indicators Based on Cohort Analysis 41
Basic Concept 41
Efficiency and Other Educational Indicators 42
Completion or Graduation Rate 42
"On Time" Graduates 42
Student/Years per Graduate 43
Input-output Ratio 43
Years Spent in School 43

Indicators Based on Enrollment Data 43
Gross Enrollment Ratio 43
Net Enrollment Ratio 43

Estimation of Transition Rates 44
Standard Method 44
Schiefelbein Method - Description 45
Schiefelbein Method - Example of Calculations 49

Projections of Resource Requirements and Costs 50

Projecting Population 52

Appendix to Chapter 4 55
Glossary of Variables and Symbols 55

CHAPTER 1. INTRODUCTION TO THE STEP SYSTEM

Introduction to the STEP System

The System for Tracking Educational Progress (STEP)[1] is designed to allow education planners and policy makers to keep track of and to project a set of basic educational indicators of an education system with an emphasis on the primary level. With it, based on a time series of enrollment data, and population data, the user can:

- estimate the transition rates within a primary school system;
- project those rates into the future using various approaches;
- project annual population by single-year age groups;
- project student enrollment into the future;
- project resource requirements, teacher demand and supply, and recurrent and capital costs; and
- calculate various educational efficiency measures.

The projection methodologies used in STEP have been used for several years in developing country settings. Those methodologies are described in more detail under the Technical Description section of this document. In this section we describe the implementation of the system on the microcomputer and how to use it.

STEP Modules

STEP uses a modular approach. This gives the user the flexibility to run modules independently and/or to concentrate on those modules that are of particular interest. Each of the six modules of STEP corresponds to a set of related tasks. The module ESTIMATE estimates the historical education transition rates for entry into the first grade of primary school, the repetition rates and the drop out rates; the module PROJRATE projects those rates forward for a specified time period; the module EDPROJ projects school enrollment; the module COHORT estimates various educational efficiency measures; the module POP allows the user to project annual population by sex and single-year age groups; and the module COST projects teacher supply and demand by level of experience and qualification, learning resource requirements, construction costs, teacher salary costs, and other recurrent and capital costs.

Using STEP on the Microcomputer

Hardware Environment

Because of the portability, access and ease of use offered by today's microcomputers, and because IBM or compatible equipment has become the de

---

1. Formerly the Primary Education Tracking System (PETS)

facto standard for social science analysis and simulation modeling, STEP has been implemented for the IBM/PC, IBM/XT or IBM/AT microcomputer and compatibles. Because of the large amount of data used by STEP and because of the software under which it is implemented the following configuration of equipment is necessary:

- IBM microcomputer or compatible;
- 256K of RAM;
- 3 MB of hard disk space;
- numeric coprocessor;
- printer (optional);
- color monitor (optional);
- graphics display adapter (optional).

Software Environment

STEP uses and produces a large amount of information and data. Two different software approaches were tried during the initial stages of development of STEP. The first one used the spreadsheet program Lotus 1-2-3. This has the advantage of being familiar to many microcomputer users and is fairly easy to use. It has the disadvantage of becoming awkward and even difficult to use when the application becomes complex, large, modular, or menu-driven.

The second approach, and the one that was chosen, was development of a set of Turbo PASCAL programs that runs under a software "shell" called Host. Host was specifically developed for modeling applications such as STEP. It was developed by the Research Triangle Institute with funds provided by USAID through the Integrated Population Development Planning (INPLAN) project and the BRIDGES project. It is as easy to use as Lotus 1-2-3, can manipulate large amounts of data, and can project faster than Lotus 1-2-3. It also has several features that make properly implemented models extremely flexible and adaptable to a wide variety of settings. An introduction to the facilities of Host, that it is necessary to know to use STEP, is given in a separate manual called "STEP by Steps".

Organization of this Manual

The rest of this manual is organized in three sections. Chapter 2 describes the use of the STEP modules. In Chapter 3, how STEP can be applied to a new application is described. Chapter 4 presents a complete technical description of the methodology underlying STEP.

CHAPTER 2. USING STEP

In running STEP there are three fundamental tasks that the user performs with the help of Host: data entry and editing, running modules, and viewing results. Each of these tasks can be done by different facilities available under Host. Thus the Host EDIT facility allows the user to enter data values into the data base and/or to change existing values. The Host RUN facility allows the user to execute a STEP module for a specified number of periods. The Host TABULATE facility allows one to view variables in the system in predefined tabular formats and to display those tables on the screen, the printer or to a disk file. Alternatively, the Host GRAPH facility allows one to view the variables in predefined graphs. Each of these facilities is described in more detail in the Host User's Guide and in the STEP by Steps Manual.

STEP has over thirty variables. Some of these variables or parameters must be supplied by the user, and others may be calculated by a STEP module as an output. Some outputs of one module may be used as inputs by another module. For example, repetition rates (REPEATRT) are calculated in ESTIMATE, projected forward in PROJRATE, and used as input to project enrollment in modules EDPROJ and COHORT. The population is projected in POP and used as an input in ESTIMATE and in EDPROJ. Enrollment is projected in EDPROJ and used as input in COST.

Interactive vs. Batch Modes

The STEP system can be run either interactively or in "batch" mode. In the interactive mode, the computer queries the user as to choices to be made in estimating key parameters, methods to be used for projecting them, etc. For example, the user might be queried as to whether he or she wants to use a linear or a logistic curve to project future repetition rates. This interactive mode is very useful the first few times one runs the STEP system. After one gets some experience, however, it becomes tedious to have to answer all these questions every time one runs the system. STEP will remember the answers given the previous time the system was used, and will give them back as default answers to each question during the current run.

In batch mode, the system runs automatically and reads the answers to the questions from the data file. They will have been stored the previous time the system was run in the variable called "BATCHOPT", and can be modified manually at any time using the editor. In batch mode no questions are asked, so the user should be sure that the settings contained in variable BATCHOPT are correct. The interactive mode is also useful if one wants to change a few of the answers. In this case, one simply hits ENTER in order to give the default (remembered) answer for all questions whose answers do not change, and enters a new value for the ones to be changed. Default answers are presented

in square brackets []. The use of Batch mode is described in more detail toward the end of this chapter in the section entitled, "Running the STEP System in Batch Mode".

In addition to the variable BATCHOPT, the variables POPFILE and EDFILE contain information regarding the transfer of population and enrollment data respectively from on STEP data file to another. These variables function similarly to BATCHOPT in interactive and batch mode.

File Management Under STEP

The STEP system uses two separate Host data bases. These are linked together and information is passed between them. A single data base would have been too large for most systems and would therefore be impractical.

While the Host software provides a facility for backing up and deleting data files, the STEP user might find it helpful to use DOS commands for file management. It is likely that the user may accumulate a number of data files and that he/she may wish to store these on floppy disks rather than on the hard disk. In this case the user should use the DOS "COPY" command to copy the files to a floppy disk and then after they are safely copied, use the DOS commands "DEL" or "ERASE" to delete the files from the hard disk.

Host data files are actually six inter-linked DOS files with the same name but different extensions. Therefore the use of the DOS "wild-card" character is recommended. In the following example a set of Host data files with the name FILE1, corresponding to a projection database called FILE1, is transferred to a diskette and then erased. We assume that the user is in the data subdirectory and that the target diskette is in drive A. To copy the files to the floppy disk, use

```
C>copy file1.* a:[Enter]
```

or for added safety, use this command with the "verify" option:

```
C>copy file1.* a: /v[Enter]
```

To erase from the hard disk, use

```
C>del file1.*[Enter]
```

Make sure that you have indeed copied to the floppy disk before you erase from the hard disk.

Eventually, you may wish to copy the files back from the floppy disk to the hard disk. Assuming that you are in the sub-directory which contains the data files, and that the diskette containing the archived files is in drive A, use

```
C>copy a:file1.*[Enter]
```

There are two other sets of files that should be backed up occasionally, but not necessarily as frequently as the data files. These are the LABEL files, the TABLE files and the GRAPH files. To manipulate these, follow the

same sets of commands as above, substituting "table", "label" and "graph" for "file1". Note that data files such as FILE1, FILE2, etc., share the same LABEL, TABLE and GRAPH files, so if you intend to continue working with the system, but with new data files (such as FILE2), do not delete your TABLE, LABEL and GRAPH files.

All STEP variables (input and output) are stored on disk in a file, called a projection file under the Host system. One of the most powerful features of the Host implemented STEP system is that it is possible to save a set of assumptions in a file and the results of a projection based on those assumptions. For example, the user could project the various education transition rates using a logistic curve in PROJRATE and then project school enrollment using EDPROJ. These rates and the resulting enrollment projections might be stored in a file called, for example, FILE1. Then the user might wish to project the rates using a linear trend in module PROJRATE and subsequently re-run module EDPROJ. In this case the user could store the second set of rates and the results in a second file called, for example, FILE2. The assumptions and results that were stored in FILE1 would still exist and it is even possible to produce tables which compare the two files.

Because POP uses so much data, it has its own database which is separate from the primary STEP database which is used by ESTIMATE, PROJRATE, EDPROJ, COHORT and COST. When ESTIMATE or EDPROJ, which use population data, are run, the user is asked (if in interactive mode), whether he or she wants to use the population data from a POP database, or population data already in the STEP database. If the user chooses to use data from a POP database, the computer will ask for the name of the POP database. So it is important that when the user intends to take population from a POP database, he or she knows the name of this database in advance. If these modules are run in the batch mode, the answers to these questions must be put into the batch options variable. (See the last section of this chapter, entitled "Running the STEP System in Batch Mode".)

The six modules that make up STEP are briefly described below and information is provided on their use. A description of the methodologies underlying these modules can be found in the Technical Description section. Their use is described in this section.

STEP Modules: POP

Description of the POP Module.

POP is a population projection module which provides the user, by default, with single-year age group population data on an annual basis. It can be used as a stand-alone population projection package, because it strikes a balance between ease of use, completeness of options, and completeness of capabilities. In the context of the STEP system, it is designed to provide the modules ESTIMATE and EDPROJ with the school age population.

The INPUT data for POP are the following:

INPUT data for POP module	
ASFR	Age specific fertility rates. Need to be specified only if User-Supplied ASFR option is chosen with the variable FERTOP.
EXTMIG	External migration, in absolute numbers.
FERTOPT	Fertility options.
FEMINDEX	Index for women.
LEXP	Life expectancy.
LT	Life table L values. Need to be specified only if user-supplied option is chosen in MORTOPT.
MIGRATE	Internal migration rates.
MODFERT	Model fertility schedules. Need to be specified only if user-supplied pattern is chosen with FERTOPT.
MORTOPT	Mortality options.
POP	Base-year single-age group population. Can be input directly or, if single-year data are not available, five-year age group data can be input into POP5, and then the program INT5T01 can be run to produce the variable POP.
SEXRATIO	Ratio of boys to girls at birth.
TFR	Total fertility rate. Needs to be specified only if User-Supplied ASFR was not chose with the variable FERTOPT.

The only output variable produced by POP which is used by other STEP modules is:

OUTPUT data from POP module	
POP	annual, single-year age group population data

However, other standard demographic indicators such as the crude birth and death rates, and the population growth rates are also reported.

Use of the POP module

Specifying rates and initial values

Fertility. There are two basic ways to specify fertility:

- enter projected age-specific fertility rates if available;
- enter projected total fertility rates and choose an age-pattern of fertility.

If the user chooses to enter total fertility rates, then an age pattern of fertility must be chosen. Currently there are three model patterns available, for Sub-Saharan Africa, Asia, and Arab countries, and there is room for a user-supplied pattern.

The fertility options are chosen by editing the variable FERTOPT. The choices presented are:

User ASFR;
Sub-Saharan Mod;
Arab Model;
Asian Model;
User Pattern.

The user chooses one of these options by entering a 1 by it and leaving zeroes elsewhere. If the user chooses one of the four patterns (Sub-Saharan Model, Arab Model, Asian Model, or User Pattern) POP assumes that the user has input a total fertility rate. In this case, the program uses the total fertility rate and the chosen pattern to produce the age-specific-rates as an output. If the user chooses the first option (User ASFR) then the program reads the age-specific rates as input and produces the total fertility rate as output.

After having told the program which option he/she has chosen, the user must input the actual rates, either by editing the variable ASFR (if he chose the User ASFR option of FERTOPT), or the variable TFR (if he chose one of the four pattern options in FERTOPT). These data must be input for every period the user wants to project.

Finally, if the user has chosen to use his/her own fertility pattern, this pattern must be entered into the variable MODFERT. It is assumed by POP that the pattern will be expressed as a series of percentages (adding to 100) of births occurring in each fertile age group (from 15-19 to 45-49), each corresponding to a total fertility rate. A maximum of six series of proportions is allowed, to correspond to six total fertility rates.

Specifying mortality

Mortality is specified by entering the projected life expectancies in the variable LEXP. They must be entered for every projection year. The user can choose which life table to use, however. The choice is made through the variable MORTOPT. Editing this variable will present the user with the choices of:

C-D North;
C-D South;
C-D West;
C-D East;
UN General;
UN Lat. America;
UN Far East;
UN So. Asia;
User-Supplied.

The user must specify one of these options by entering a 1 next to it, and zeroes elsewhere. The first four options represent the Coale-Demeny tables, the second four represent the UN tables, and the last option is for a user supplied set of L columns from a life table for a particular country, if available.

The standard life table data are not included as part of the Host database for POP. Instead, they are provided as binary DOS files. If one of the first eight options is chosen, the user should make sure that the relevant file is included in the directory where POP assumes the data files are.

If the user chooses to use his/her own life table L values, these must be entered by editing the variable LT for both males and females. (Currently these values must be entered by hand. A supporting program is being written to import these data from a Lotus 1-2-3 print file or any ASCII file corresponding to a certain format.) Note that the first row of the variable LT must contain the life expectancy at age zero to which the column of L values corresponds. Also note that the first age group is 0-4, not 0. The 0-4 group data for L is just the sum of the rows labeled 0 and 1 in most life tables.

Specifying International and Internal Migration

International migration data are entered into the variable EXTMIG. These data are entered as absolute numbers, negative if there is net out-migration, positive otherwise. The data must be entered for every projection period.

Internal migration is handled automatically when the current level of the dimension REGION is set equal to two. Model migration functions are used to calculate the internal migration rate (IMR) and the number of migrants by region and by age (INTMIG).

Base Year Population

The user must enter the base year population. There are two ways to do this. If single-year age group data are available and considered reliable, these can be entered into the first year of the variable POP, for both sexes. If the single-year data are unavailable or unreliable, then five-year group data can be entered into the variable POP5. In order to provide POP with a base-year single-year group population data, the support module INT5T01 must be run, for one period only.

INT5T01 is a general purpose age decomposition routine which uses "multipliers" to decompose variables given in five year age groups into single year age groups. The inputs for INT5T01 are given in the table below. Since most users of STEP will be using INT5T01 to decompose the population, the values of the control variables INT1NAME and INT5NAME should be set as follows:

INT1NAME : POP
INT5NAME : POP5

Also, the values of variable INT5OPT should be left at 1.

INPUT Data for INT5TO1	
POP5	population data in five year age groups
INT5OPT	options for population interpolation
INT1NAME	name of output variable for INT5TO1
INT5NAME	name of INPUT variable for INT5TO1

OUTPUT Data for INT5TO1	
POP	base year population in single year age groups

The base year for the population projection need not correspond to that of the base year in the STEP data base. That is, if the STEP data base has 1980 as the first year for which historic data are entered, the base year for the population data base may be 1980 or earlier. If a period later than 1980 is specified for the population data base Host/STEP will issue a fatal error message and the program will stop.

Sex Ratio

The sex ratio at birth must be specified by editing the variable SEXRATIO. Populate assumes this will not change over the projection period, and so this need be entered only once.

Index of Females

POP needs to know, for its internal calculations, whether all the variables will be arrayed in such a way that men will be listed first, and then women, or vice-versa. If men will always be listed first, the user needs to set the variable FEMINDEX to 2, and to 1 in the opposite case.

STEP Modules: ESTIMATE

Description of the ESTIMATE Module

The ESTIMATE module estimates repetition, dropout, and enrollment rates, using either the standard method based on reported repeater data or the Schiefelbein method. For this module the INPUT data are:

INPUT data for ESTIMATE module	
Transferred from POP or input by the user:	
POP	population by single-year age groups for the historic period (and for as many periods in the future as the user may wish to project)
Input by the user:	
HISTORIC	indicates whether a given year is historic (represented by an "h") or projected (represented by a "p")
ENTAGE	age of entrance to school system
The following information is necessary only when using the "Schiefelbein" method, or if age-specific enrollment rates for the historic period are desired:	
AGGENROL	enrollment in the aggregate (last) age group by grade
ENROLAGE	enrollment by grade and single-year age groups
The following information is necessary only when using the "standard" method:	
NREPEAT	number of repeaters by grade
ENROL	enrollment by grade - can be aggregated automatically from the variable ENROLAGE if data have already been entered there

The use of the variable ENROLAGE deserves special attention. It is important for the Schiefelbein method that the lowest age group in the first grade contains a number of students that is much lower than the next higher age group. Thus, for example, if the lowest age group is "6 and under" and contains 60,000 students, and the next is "7" and contains 70,000, as would be typical of many school systems, then the Schiefelbein method will not work properly. (The main consequence will be that the estimated rates of entrance into the school system from the population at large will be too low.) There are two ways to deal with this situation. First, the data in the "6 and under" category could be disaggregated into two or three categories: "6" and "5 and under", or "6", "5" and "4 and under" until one gets a category with a very small number in it (100 or 1000 times less than the next higher one). This first alternative is relatively difficult because it is not clear where to obtain this disaggregation. If the school system had the data presumably they could provide it. The second, easier alternative is to simply put in a category lower than the lowest one for which one has data, and to put only zeroes in it. This is the alternative we have used in the sample data in this document.

When the user chooses the Schiefelbein method, some additional information is required about the repetition rate in the last (usually 6th) grade, the number of dropouts, and the age distribution of enrollment. See below.

The primary OUTPUTS of this module are:

OUTPUT data from ESTIMATE module	
REPEATRT	repetition rate by grade
DROPRT	dropout rate by grade
ENTRT	enrollment rate of new students in first grade
ASER	age-specific enrollment rates
ENROL	enrollment by grade and sex
PROMRT	promotion rate by grade

Note that the outputs of this module are calculated only for the historic period. When using the Schiefelbein method, the estimated historical repetition rate for the sixth grade is simply the one assumed by the reader and entered into the computer at run-time. Both with the Schiefelbein and the standard methods, the estimated sixth grade dropout rate will be zero by assumption. Note that with the standard method, the estimated dropout rates will be meaningless for any year followed by a year with no enrollment data.

Use of the ESTIMATE Module

While executing ESTIMATE from the RUN facility of Host, the user must specify, as with any module, the number of periods to execute the module. In ESTIMATE, the user should specify a number that is equal to the number of historic periods. If you cannot recall the length of the historic period, specify one that is AT LEAST as large as the number of historic periods. That is, if there are five historic periods, the user should specify at least FIVE. Since no actual "projections" take place in this module, any periods beyond the historic period are ignored.

The last year of historic enrollment data should be the last year for which the indicator variable HISTORIC is set to "h", both with the Schiefelbein and the standard method.

When the ESTIMATE module is executed, as with all modules, the user is provided with an opening screen and with a screen that briefly describes the module.

Then he/she is asked whether the module is to be run interactively or in batch mode. If the module is to be run in batch mode, the user must have already entered his/her choices in the BATCHOPT and POPFILE variables, or else the values already there from the previous run will be used. (See section entitled "Running the STEP System in Batch Mode" at the end of this chapter.)

If the user chooses interactive mode a series of questions must be answered. Next, the user is asked whether he/she wishes to use "Standard method" which uses official data, or whether he/she wishes to use the method developed by Schiefelbein.

Since this module requires population data, the user will be asked whether he/she wishes to use the data already in the STEP database, or to bring in data generated by the POP module. If the user wishes to bring in data from the POP module, he/she will need to enter the name of the database containing this data, so it is important to recall this name before running the module. Also, when importing data from a POP database, the first period specified for the ESTIMATE module may not precede the first period in the POP database.

For the standard method, the user must have entered data on the number of repeaters through the Host EDIT facility. Such data are entered into the variable NREPEAT.

For the Schiefelbein method, other information must be entered:

First, the user is prompted to enter an estimate of the 6th grade repetition rate. Here, the user would enter his/her best guess of this rate. All other repetition and dropout rates are calculated based on that rate.

Second, the user is prompted for an assumption about the percentage of non-promoted students who are dropouts. In other words, if 90 out of 100 students in grade 1 pass to grade 2, the user must supply a percentage of the 10 students who do not pass who dropout. If the user puts that percentage at 50% then, in this example, 5 students would drop out and the other 5 would repeat first grade. In this case the user would enter 50.

Third, some additional information is required about the disaggregation of enrollment data. In many countries it is possible that there are some students, particularly in the later grades, who are "over-age". For example it is not unusual to have a number of 17 or 18 year olds in the sixth grade. enrollment data by age and grade that are published in official sources often report the number of such over-age students, but the breakdown of their ages is often not sufficiently detailed for the Schiefelbein method to work. For example, the last reported age group may be 14, rather than 18. The aggregated data for the last age group of the ENROLAGE matrix must be entered initially into the variable AGGENROL. When these aggregated data are present the user must tell the computer to disaggregate the data when prompted.

If disaggregation is necessary, the user is prompted to specify the last age group. It is important to enter the correct value here, otherwise the disaggregation will be incorrect, as will all the estimated rates.

For the standard method, the computer requests information on whether the enrollment data were entered in disaggregated form (through the variable ENROLAGE) or directly in aggregated form (through the variable ENROL). If the data were entered in disaggregated form, the computer asks whether the last age group is actually an aggregated one, as in the Schiefelbein method.

Finally, after the estimation of the transition rates is completed, the computer asks the user whether age-specific enrollment rates are desired. These rates have meaning only if age-specific enrollment data were entered. Note that there is no need to re-do these calculations unless new data on population or enrollment by grade and age have been entered.

STEP Modules: PROJRATE

Description of the PROJRATE Module

Future dropout, repetition, and first grade enrollment rates are necessary inputs for the calculation of enrollment projections. The user has the choice of entering his/her own scenario about the evolution of these rates, or he/she can let the computer do so by running PROJRATE. The module PROJRATE projects repetition, dropout, and enrollment rates for a specified number of periods, using one of five approaches that can be chosen by the user:

- future rates fixed at the level of the last historic value
- future rates fixed at the level of the historic mean
- varying future rates that follow a logistic curve
- varying future rates that follow a linear trend
- varying future rates that attain target levels at a target date, where the targets and the target year are given by the user

The INPUT data for this module are the historic values of:

INPUT data for PROJRATE module	
REPEATRT	repeater rates by grade
D .PRT	dropout rates by grade
ENTRT	entrance rate for the first grade
TARREP	target repetition rate - necessary only for target setting projections
TARDROP	target dropout rate - necessary only for target setting projections

Before running PROJRATE, the historic values of these rates must have already been estimated using the module ESTIMATE, or they could have been entered manually from some other source.

The module also requires as INPUT information about which periods to use as the basis for the projections:

	Indicator variables for PROJRATE
HISTORIC	signals which periods are historic, which are to be projected
INCLUDE	signals which of the historic periods to be included as the basis for the projections - setting INCLUDE equal to "n" (no) will result in the data for that year being ignored in estimating projection trends

The variable INCLUDE is useful because it allows the user to ignore periods which may be considered atypical and so may not want to be included in the calculation of future trends. Note that any period designated as **not** included will be subsequently ignored in the module COHORT.

The variable HISTORIC should remain set as it was for the ESTIMATE module. If the Schiefelbein method was used for estimating historical rates, the variable INCLUDE should be set so that the last year for which there are enrollment data IS NOT included. If the standard method was used, then the variable INCLUDE should be set so that the last year with data on enrollment and numbers of repeaters IS NOT included. This is due to the fact that, with the Schiefelbein method, to estimate the rates for any given year, enrollment data for that year **and the next** are needed. With the standard method, enrollment data for the given year and enrollment and repeater data for the next are needed. For example, if the last year with enrollment data is 1984, the estimated rates for 1984 itself are meaningless. Therefore, 1984 must not be included in estimating the trends - in fact, the rates for 1984 must themselves be projected. NOTE: this applies also to any gaps in the data. Thus, for example, if there are enrollment data for 1974-1976, and 1981-1984, then even though 1974-1984 are designated "h" with the HISTORIC variable, both 1976 and 1984 must be designated "n" (for "no") with the INCLUDE variable: since there were no enrollment data in 1980 or 1985, the estimated rates for 1979 and 1984 are meaningless. Of course, 1977-1980 and 1985 and onward must also be designated "n" with the INCLUDE variable. Here is the rule: if you have n historic years WITH DATA, but with m gaps in the data (regardless of their length), then the number of included years can be AT MOST n-m-1, making sure that the last historic year is not included. All this should be checked carefully. Study the following table to make sure you understand it.

Example of use of HISTORIC and INCLUDE			
Year	Have historic data?	HISTORIC	INCLUDE
1974	Yes.	h	y
1975	yes.	h	y
1976	yes.	h	n
1977	No.	h	n
1978	No.	h	n
1979	No.	h	n
1980	No.	h	n
1981	Yes.	h	y
1982	Yes.	h	y
1983	Yes.	h	y
1984	Yes.	h	n
1985	No.	p	n
1986	No.	p	n
1987	No.	p	n
1988	No.	p	n
1989	No.	p	n
1990	No.	p	n

Note: there are 7 historic years with data, there is 1 gap in the data, therefore there are 7-1-1=5 years with INCLUDE set to "y".

Of course, all of the above applies also to years in which the estimated rates are considered to be "atypical". In these cases, the year with the "atypical" rate should not be included. If it is the enrollment data that are considered atypical, then the year with atypical enrollment data AND THE PREVIOUS ONE should not be included.

The OUTPUTS of this module are the projected values of:

OUTPUT data from PROJRATE module	
REPEATRT	repeater rates by grade
DROPRT	dropout rates by grade
ENTRT	enrollment rate for the first grade
PROMRT	promotion rate by grade
TREP	t-statistic for the slope of the repetition regression
TDROP	t-statistic for the slope of the dropout regression

Use of the PROJRATE Module

Again the user will execute PROJRATE from the RUN facility of Host and, after selecting a "projection" or data file, will then be prompted to specify the number of periods to project. In PROJRATE, as with EDPROJ, the number of

periods that are given here include the number of historic periods. Thus if the user has an historic period of data corresponding to 1979 - 1984, and wants to project until the year 1994, he/she would specify a projection period of 15 (1994-1979=15). If the user attempts to project for too short a period the PROJRATE module will detect this and will so inform the user.

Then he/she is asked whether the module is to be run interactively or in batch mode. If the module is to be run in batch mode, the user must have already entered his/her choices in the BATCHOPT variable, or else the values already there from the previous run will be used. (See section entitled "Running the STEP System in Batch Mode" at the end of this chapter.)

When in interactive mode, once the projection period has been specified and the user executes the module, he/she will subsequently be asked to choose the method wanted for projecting the rates. An initial choice is made between fixed and varying rates. Then a choice is made among either the two fixed methods or the three available varying methods, depending on whether the user chose fixed or varying rates.

After projecting the rates, the user should check the t-statistics for the trend estimates. Projections based on weak trends are very risky, and checking this statistic is the minimum that should be done by the careful user.

For all choices except the target method, there are no further decisions required of the user and the program proceeds.

If the user has chosen to project using target rates, he/she is reminded that these targets, that are in the variables TARREP and TARDROP, must be up to date. Targets can be updated through the Host EDIT facility. If the target variables are up to date, the user then specifies the date at which the targets are to be achieved. Obviously, this date must be within the projection period.

STEP Modules: EDPROJ

Description of the EDPROJ Module

The module EDPROJ projects enrollment in primary school for a specified number of periods using the standard grade transition method of projecting education enrollment. For this module the INPUT data are:

INPUT data for EDPROJ module	
Transferred from POP or input by the user:	
POP	population by age
REPEATRT	repeater rates by grade
DROPRT	dropout rates by grade
ENTRT	entrance rate for the first grade
ENDAGE	oldest "legal" school age
ENROL	base year enrollment by grade

The rates must have already been supplied by running the other modules in the system, i.e. ESTIMATE and PROJRATES, or they may have been entered manually.

The OUTPUTS of this module are:

	OUTPUT data from EDPROJ module
ENROL	enrollment by grade
GER	gross enrollment rate
TOTENROL	total (aggregated across grades) enrollment in primary school
SCHLPOP	total (aggregated across ages) population of school age

Use of the EDPROJ Module

As with other modules, the user of EDPROJ will have to tell Host the number of periods to project. Here, there must be input data for each of the projection years. Hence if the user has an historic period of 1979 - 1984 and wants to project until 1994 he would specify a run of 15 periods, 10 of which would be projections. Here, the module PROJRATE would have had to have been run for 15 periods. In addition, population data for the projection period as well as the historic period would have to be present.

Then he/she is asked whether the module is to be run interactively or in batch mode. If the module is to be run in batch mode, the user must have already entered his/her choices in the BATCHOPT and POPFILE variables, or else the values already there from the previous run will be used. (See section entitled "Running the STEP System in Batch Mode" at the end of this chapter.)

Since this module requires population data, the user will be asked whether he/she wishes to use the data already in the STEP database, or to bring in data generated by the POP module. If the user wishes to bring in data from the POP module, he/she will need to enter the name of the database containing this data, so it is important to recall this name before running the module. As with the ESTIMATE module, if data are being imported from a POP database, the first period specified when running EDPROJ must not precede the first period in the POP database.

Once this module is executed, no further input is required on the part of the user.

STEP Modules: COST

Description of the COST module

The COST module allows the user to forecast

- teacher demand by salary grade
- demand for administrators by salary grade
- other recurrent costs
- capital costs such as classroom construction, land acquisition etc., by type of resource

As currently configured, the COST module is sufficient only for projection of fairly general cost categories. Detailed cost and resource requirement projections for school systems is complex enough to require specialized software. This software is being developed under the BRIDGES project, but is not yet available.

The input data for the COST module include:

INPUT data for COST module	
TOTENROL	Total enrollment (summed across grades and ages) by sex. Can be input manually or result from EDPROJ module.
SYGRAD2	Student/years per graduate, accounting for dropouts. This variable is an output of the COHORT module. See section on Efficiency and Other Educational Indicators of the Technical Description chapter.
UNITCOST	Cost per unit of each resource, by cost or salary grade level. Needs to be entered for BASE year only.
COSTINCR	Rate of increase for UNITCOST variable.
COSTTYPE	Matix showing the uses and descriptions of cost variables classified by cost or salary grade and by type of cost category. This serves only as an aid in using the variables UNITCOST, QUALDIST, and RES_COEF.
QUALDIST	Distribution of resources by cost or salary grade level. For the base year, the current distribution must be entered. For the target year, the desired distribution must be entered.

INPUT data for COST module (continued)	
RES_COEF	A matrix detailing resource use coefficients and various other indicators. The rows correspond to different types of resources, such as teachers, classrooms, etc. The first two columns show the base year and target year number of students per unit of the resource, such as students per teacher, etc. If the concept is not relevant to the resource, it must be set to 1. The second two columns show the base year and target year number of shifts per unit of the resource, such as number of shifts per teacher, etc. Again, if the concept is irrelevant, these must be set to 1. The fifth column contains the minimum level of resource use. This is relevant for central resources such as administrators, etc., for which there is a minimum level that does not depend on the number of students in the system. If the concept is irrelevant for a particular resource, it must be set to 0. The sixth column indicates whether the resource is a capital good. If the resource is a capital good, a 1 must be entered, otherwise a 0 must be entered. Finally, the seventh column indicates what proportion of total costs the "Other Costs" category represents. For example, if "Other Costs" are 5% of Total Costs, a 0.05 needs to be entered here. Note that this must be entered as a proportion, not as a percentage (e.g., 0.05, not 5). For all resources except "Other", the "Prop." column must have 0s. For the "Other" resources category, both the students-per-unit-of-resource column and the "Shifts" column must have only 0s.

OUTPUT Data for COST Module	
RES_COST	Resource cost by type
RES_USE	Resource use in physical units by type
COST_CAP	Per student costs
COST_GRA	Cost per graduate

Use of the COST Module

The COST module may be run independently of the other modules, as long as the data that are normally output by other modules (the variables TOTENROL and SYGRAD2) are entered manually. If it is run as part of STEP, it should be run after EDPROJ and COHORT have been run, so that the outputs from those modules can be used. Evidently, EDPROJ and COHORT need to be re-run prior to running the COST module only if something has been altered that would change the

projections of enrollment or of internal efficiency indicators. If the changes made affect only categories such as unit cost, or per student resource requirements, then COST can be re-run without re-running EDPROJ or COHORT.

STEP Modules: COHORT

Description of the COHORT Module

The module COHORT projects the history of a cohort of 1000 entrants in primary school according to given repeater and drop out rates. The module allows the user to use two alternative methods:

- fixed drop out and repeater rates;
- varying or contemporaneous dropout and repeater rates.

The required INPUTS of the COHORT module are:

INPUT data for COHORT module	
REPEATRT	repeater rates by grade
DROPRT	drop out rates by grade

As with the EDPROJ module, these rates must have already been estimated by running other modules in the STEP system, or may have been entered manually.

The OUTPUTS of the COHORT module are:

OUTPUT data from COHORT module	
IORATE	the input/output ratio
GRADRATE	number who graduate out of 1000 entrants
YRSSCHL	average years spent in school
SYGRAD	student/years per graduate not counting student/years "wasted" by drop-outs
SYGRAD2	student/years per graduate, counting student/years "wasted" by drop-outs
ONTIME	number of graduates who graduate "on time", i.e. have not repeated

Use of the COHORT Module

Use of the COHORT module is similar to that of module EDPROJ. An appropriate number of periods must be specified and inputs present for the length of the projection period.

Then he/she is asked whether the module is to be run interactively or in batch mode. If the module is to be run in batch mode, the user must have already entered his/her choices in the BATCHOPT variable, or else the values already there from the previous run will be used. (See section entitled "Running the STEP System in Batch Mode" at the end of this chapter.)

Once the model is run the user is asked to choose between two alternative methods.

In the first method, all transition rates are assumed constant and equal to those during the entry year. For this method there must be as many periods of data for the transition rates as the current number of periods for which COHORT is being executed. This method is applied to the historic period for periods which are "included" by the INCLUDE variable being set to "y". It is also applied to the whole projection period.

In the second method, the transition rates that are applied change according to the actual calendar year in the future school history of the cohort of entrants. Experience with STEP has shown that it may take as long as 15 years for a cohort of entrants to finally graduate or dropout. In this case, a ten year projection would require AT LEAST twenty five years of future rates. For this reason, the program reminds the user who has chosen method 2 that a sufficient amount of data for future repeater and dropout rates must be in the data base. The user has the option of exiting the program and re-running PROJRATE if he/she determines that there are not enough periods of transition rate data.

Because this second method breaks down if there are gaps in the data, and since gaps in the data can only occur in the historic period, it cannot be applied in any periods designated "h" with the HISTORIC variable. The experienced user can trick the system by designating even the historic periods as "p", but this is fairly dangerous, since one might forget that this was done, and run into trouble later, when re-estimating or re-projecting the transition rates. Also, this should be done only if one is absolutely sure of the validity of the historical data, and if there are no gaps in them.

Note that the module COHORT can be used for analyzing the impact of programs oriented to the first few grades. For example, one can calculate the "fourth grade completion rate" by forcing the computer program to think there are only 4 grades right before running the module COHORT. This requires EXTREME caution, because it is easy to forget one has done this, and it will bias any subsequent exercises with the ESTIMATE, PROJRATE, and EDPROJ modules. The way to do this is as follows. Use the D)efine function in the Main Menu, then the D)imension function in the Define Menu, and then the E)dit function in the Dimension Menu. Enter the name of the GRADE dimension. Then use the cursor to go to the cell for the Current No. of Levels, and input the number 4, and press RETURN. Then press [s] for Save, [Enter] to save the dimension under its own name, and press [ESC] several times until you get back to the main menu. After running the module COHORT, go back and change the current number of levels of the grade dimension to what it was.

Running the STEP System in Batch Mode

In order to run the system in Batch mode, the user needs to be satisfied that all the options that were used last time (e.g., linear vs. logistic projection of rates, Schiefelbein vs. standard estimation method) are still valid. If not, the user can edit the variable BATCHOPT. This variable contains a record of the choices that were made the last time the system was used. Any of these choices can be changed by editing this variable. The list of choices appears as follows:

BATCHOPT Variable
ESTIMATE Method (1=Std, 2=Schief)
Default Grade 6 Rep. Rte MF
Default Grade 6 Rep. Rte Male
Default Grade 6 Rep. Rte Female
Drop out proportion Male, Female
Drop out proportion Male
Drop out proportion Female
Last age disagg.? (1=no, 2=yes)
Aggregated Age
Disagg. by Age (1=disagg., 2=agg.)
Age Spec. Enrol. Rte? (1=yes, 2=no)
PROJRATE Meth. (1=Varying, 2=Fix)
Varying Meth. (1=log, 2=lin, 3=tar)
Target year for Target Rates
Fixed Rates val. (1=mean, 2=last)
COHORT Meth (1=fixed, 2=varying)
Target Year for COST Module

The meaning of each choice is as follows:

Estimate Method (1=Std, 2=Schief)

Whether to use the Schiefelbein or standard method in estimating transition rates. 1=standard, 2=Schiefelbein.

Default Grade 6 Rep. Rte MF

Assumed 6th grade repetition rate to be used if data are not available and are not disaggregated by sex. Must be entered as a %, i.e., 4, not 0.04.

Default Grade 6 Rep. Rte Male

Assumed 6th grade repetition rate to be used for males if data are not available and are disaggregated by sex. Must be entered as a %, i.e., 4, not 0.04.

Default Grade 6 Rep. Rte Female

Assumed 6th grade repetition rate to be used for males if data are not available and are disaggregated by sex. Must be entered as a %, i.e., 4, not 0.04.

Drop out proportion Male, Female

Proportion of non-promoted who will be assumed to be dropouts when data are not disaggregated by sex. NOT a dropout rate. Must be entered as a %, e.g., 4, not .04.

Drop out proportion Male

Proportion of male non-promoted who will be assumed to be dropouts when data are disaggregated by sex. NOT a dropout rate. Must be entered as a %, e.g., 4, not .04.

Drop out proportion Female

Proportion of female non-promoted who will be assumed to be dropouts when data are disaggregated by sex. NOT a dropout rate. Must be entered as a %, e.g., 4, not .04.

Last age disagg.? (1=no, 2=yes)

Whether the last age group in the age-grade enrollment data was an open-ended (e.g., 14 and older) or closed (e.g., just 14) age group.

Aggregated Age

If the variable above was open-ended, the open-ended age group needs to be entered here (e.g., 14).

Disagg. by Age (1=disagg., 2=agg.)

If the standard method is chosen for estimating the transition rates, the computer must be told whether the enrollment data were entered disaggregated by age (in the ENROLAGE variable) or aggregated across ages (in the ENROL variable). 1=disaggregated, 2=aggregated.

Age Spec. Enrol. Rte? (1=yes, 2=no)

After the ESTIMATE module is finished estimating all the transition rates, it can go on to calculate age-specific enrollment rates. These can be useful in calculating the historical net enrollment ratios. But it does take some time. 1=yes, calculate; 2=not, don't calculate.

PROJRATE Meth. (1=Varying, 2=Fix)

In the PROJRATE module, the rates can be projected at a fixed level, or according to a trend. 1=vary according to a trend, 2=fixed.

Varying Meth (1=log, 2=lin, 3=tar)

If varying future rates are chosen, the trend can be either logistic, linear, or according to a target. 1=logistic, 2=linear, 3=according to a target. Remember that if the target method is chosen, the target repetition and dropout rates must be entered in the variables TARREP and TARDROP.

Target year for Target Rates

If the trend for the transition rates is projected according to a target, the target year (e.g., 2000) must be entered here.

Fixed Rates val. (1=mean, 2=last)

If the future rates are to projected at a fixed level, this level can be fixed at the historical mean, or at the value of the last historical year. 1=historical mean, 2=value of last historical year.

COHORT Meth (1=fixed, 2=varying)

When using the cohort method, either varying or fixed transition rates can be used. 1=fixed, 2=varying/contemporaneous. (This is an important but somewhat slippery concept. Refer to the sections entitled "Use of the COHORT Module" in Chapter 2, and "Efficiency Indicators Based on Cohort Analysis in Chapter 4.)

Target Year for COST Module

Target year for cost module.

CHAPTER 3. CONFIGURING STEP TO A NEW APPLICATION

As indicated earlier, one of the advantages of the **Host** modeling system under which STEP runs is that it can be adapted from one situation to another without the need to alter the original computer programs of the system. Host and STEP have been set up in such a way that the user can change a few settings in the STEP data base, enter new data and the system is ready to use. For more detailed knowledge about **Host** related fuctions, the user should consult the **Host** User's Guide. Some of the material described below is also covered in the "STEP by Step" tutorial.

In applying any **Host** model to a new situation there are several steps involved:

Step 1: Configure the data base to conform to the structure of the application by setting the dimension levels.

Step 2: Edit existing labels and define new labels which correspond to the new dimensions.

Step 3: Enter new data into the data base.

Step 4: Define new tables and graphs if desired.

In configuring the data files for the new application, existing **Host** STEP and POP files should be used. There are three possibilities. First, you can use the two "empty" files distributed with the STEP system. These Host files are called STEP0 and POP0 and contain variable definitions but no data. A second strategy is to use data bases with existing data, such as was also distributed (e.g. GUAT and POPGUAT). Of course with this approach you will need to replace existing values with new ones as appropriate. Yet a third possibility is to start with an exicting data file, "clear" the values of selected variables using **Host**'s Define/Variable/Clear facility, and then enter new variable values.

Whichever approach you use, be sure to use the **Host** data file you want when changing dimensions, etc. The easiest way to do this is to specify that file as the primary data file in **Host**'s Configure/Files facility.

Step 1: Setting the Current Maximum Levels of Dimensions

To configure the STEP and POP data bases, the first step is to set the levels of the dimensions of the variables to be equal to the correct and appropriate current levels for the application. For example, the dimension GRADE, which is used by many STEP varibles, (see Table below), has a maximum potential level for any application of twelve. For most applications, however, this would be a lower number, for example 6. This means that under the current application there would be six grades (i.e. a current maximum of six). Similar remarks apply to the other dimensions in both the STEP and POP data bases.

The tables immediately below show the dimensions for the STEP and POP data bases. To change the current maximum levels, use the Define/Dimension/Edit command of **Host**. For more information regarding dimensions, consult the Host User's Guide.

Dimensions Used by STEP Variables in the STEP Data Base

Name	Maximum Levels	Current Levels	Starting Value	Increment Between Levels	Maximum Effective Level	Default Label Format
AGE	86	76	0	1	85	AGEGP
ANSWER	40	17	1	1	40	BATCHQUE
COEF_TYP	10	7	1	1	10	COEF_TYP
COST_TYP	7	5	1	1	7	COST_TYP
ENAGE	19	19	5	1	23	ENAGE
GRADE	12	6	1	1	12	GRADE
PERIOD	1024	27	1974	1	2997	NONE
QUAL_TYP	14	10	1	1	14	QUAL_TYP
R	6	2	1	1	6	R
RES_TYPE	11	8	1	1	11	RES_TYPE
SEX	2	1	1	1	2	SEX
YEAR	40	27	1974	1	2013	NONE

Dimensions Used by POP Variables in the POP Data Base

Name	Maximum Levels	Current Levels	Starting Value	Increment Between Levels	Maximum Effective Level	Default Label Format
AGE	86	76	0	1	85	NONE
AGE5	18	16	0	1	17	AGGPOP
ASFR	7	7	1	1	7	ASFR
BEERSCOL	5	5	1	1	5	NONE
BEERSROW	25	25	1	1	25	NONE
FERTAGES	2	2	1	1	2	FERTAGES
FERTOPT	5	5	0	1	4	FERTOPT
INT5OPT	2	2	1	1	2	INT5OPT
LTCOLS	24	6	1	1	24	NONE
LTFILE	8	8	1	1	8	LTFILE
LTROWS	20	19	1	1	20	LTROWS
MFCOLS	8	8	1	1	8	MFCOLS
MFROWS	6	1	1	1	6	MFROWS
MODFERT	4	4	1	1	4	MODFERT
MORTOPT	9	9	1	1	9	LTFILE
PERIOD	1024	32	1970	1	2993	NONE
REGION	2	1	1	1	2	REGION
SEX	2	2	1	1	2	SEX

Step 2: Defining and Editing the Label Library

After the current maximum levels of the dimensions are set, the next step is to make sure that the labels which correspond to each level of a dimension are correct. For complete information about labels, once again the user should consult the Host User's Guide. In the Appendix to this chapter the labels from the STEP and POP data bases which are distributed with the system are listed. To avoid confusion, there is only one set of label libraries although there are two data bases.

In the table describing the dimensions, you will see that many of these are "attached" to dimensions. Others may be used in tables and graphs.

To alter old labels or define new ones use the Define/Labels/Edit or Define/Labels/Add commands of Host. Altered existing labels can be saved under new names so that the originals can be left intact.

Step 3: Entering New Data for the New Application

Once the structure of the data base is ready it is time to enter the data for the new application. If you have been working with the "empty" Host files, then you will need to fill in the necessary input data for each module as required. If you are using the example Host files which contain data you can either overwrite the existing values with new ones or, perhaps better, you can use Host's Define/Variable/Clear facility to clear data from the appropriate variables.

Of course, once the new base period values for the variables have been entered for the new application you will run the various modules. This will also generate new data for the new application and test to see if any non-sensical inputs have been made.

Be sure to reset the starting value of the dimension PERIOD by using Define/Dimension/Edit. Also be careful to have appropriate values in the control variables INCLUDE and HISTORIC.

Step 4: Defining New Tables and Graphs

Once the new data have been entered and the dimensions and the labels have been updated for the new application you may wish to add new graphs and tables or change the presentation of the existing ones. This can be done by using the Define/Tables/Edit, Define/Tables/Add or Define/Graph/Edit of Define/Graph/Add facilities of Host.

It should be noted that most of the tables and graphs in the distribution graph and table libraries use the default labels of the dimensions. This means that any changes to the dimensions and their corresponding labels will automatically be used by the graphs and tables.

Variables in the STEP Data Files

Name	Size	Type	Periodic	Number of Dimensions	Dimension Names	Maximum Levels
AGGENROL	192	real	Yes	2	SEX	2
					GRADE	12
ASER	1376	real	Yes	2	AGE	86
					SEX	2
BATCHOPT	80	integer	No	1	ANSWER	40
COSTINCR	616	real	No	2	RES_TYPE	11
					COST_TYP	7
COSTTYPE	19712	string	No	2	RES_TYPE	11
					COST_TYP	7
COST_CAP	16	real	Yes	1	SEX	2
COST_GRA	16	real	Yes	1	SEX	2
DROPRT	192	real	Yes	2	SEX	2
					GRADE	12
ENDAGE	2	integer	No	0	NONE	NONE
ENROL	192	real	Yes	2	SEX	2
					GRADE	12
ENROLAGE	3648	real	Yes	3	SEX	2
					GRADE	12
					ENAGE	19
ENTAGE	2	integer	No	0	NONE	NONE
ENTRT	16	real	Yes	1	SEX	2
GER	16	real	Yes	1	SEX	2
GRADRATE	16	real	Yes	1	SEX	2
HISTORIC	40	char	No	1	YEAR	40
INCLUDE	40	char	No	1	YEAR	40
IORATE	16	real	Yes	1	SEX	2
NREPEAT	192	real	Yes	2	SEX	2
					GRADE	12
ONTIME	16	real	Yes	1	SEX	2
POP	1376	real	Yes	2	AGE	86
					SEX	2
POPFILE	1536	string	No	1	R	6
PROMRT	192	real	Yes	2	SEX	2
					GRADE	12
QUALDIST	2464	real	No	3	RES_TYPE	11
					QUAL_TYP	14
					SEX	2
REPEATRT	192	real	Yes	2	SEX	2
					GRADE	12
RES_COEF	1760	real	No	3	RES_TYPE	11
					COEF_TYP	10
					SEX	2
RES_COST	176	real	Yes	2	RES_TYPE	11
					SEX	2

Variables in the STEP Data Files (continued)

Name	Size	Type	Periodic	Dimensions	Names	Levels
RES_USE	176	real	Yes	2	RES_TYPE	11
					SEX	2
SCHLPOP	16	real	Yes	1	SEX	2
SYGRAD	16	real	Yes	1	SEX	2
SYGRAD2	16	real	Yes	1	SEX	2
TARDROP	192	real	No	2	SEX	2
					GRADE	12
TARREP	192	real	No	2	SEX	2
					GRADE	12
TDROP	192	real	No	2	SEX	2
					GRADE	12
TOTENROL	16	real	Yes	1	SEX	2
TREP	192	real	No	2	SEX	2
					GRADE	12
UNITCOST	1232	real	Yes	3	RES_TYPE	11
					COST_TYP	7
					SEX	2
YRSSCHL	16	real	Yes	1	SEX	2

Variables in the POP Data Files

Name	Size	Type	Periodic	Number of Dimensions	Dimension Names	Maximum Levels
ASFR	112	real	Yes	2	ASFR	7
					REGION	2
BEERS	1000	real	No	2	BEERSROW	25
					BEERSCOL	5
BIRTHS	32	real	Yes	2	SEX	2
					REGION	2
CBR	16	real	Yes	1	REGION	2
CDR	16	real	Yes	1	REGION	2
EMR	16	real	Yes	1	REGION	2
EXTMIG	8	real	Yes	0	NONE	NONE
FEMINDEX	1	byte	No	0	NONE	NONE
FERTAGES	2	byte	No	1	FERTAGES	2
FERTOPT	5	byte	No	1	FERTOPT	5
IMR	16	real	Yes	1	REGION	2
INT1NAME	256	string	No	0	NONE	NONE
INT5NAME	256	string	No	0	NONE	NONE
INT5OPT	2	byte	No	1	INT5OPT	2
INTMIG	2752	real	Yes	3	AGE	86
					SEX	2
					REGION	2
LEXP	32	real	Yes	2	SEX	2
					REGION	2
LT	7680	real	No	3	SEX	2
					LTROWS	20
					LTCOLS	24

Variables in the POP Data Files (continued)

Name	Size	Type	Periodic	Dimensions	Names	Levels
LTFILES	2048	string	No	1	LTFILE	8
MIGRATE	16	real	Yes	1	REGION	2
MODFERT	1536	real	No	3	MODFERT	4
					MFROWS	6
					MFCOLS	8
MORTOPT	9	byte	No	1	MORTOPT	9
NATLCBR	8	real	Yes	0	NONE	NONE
NATLCDR	8	real	Yes	0	NONE	NONE
NATLEMR	8	real	Yes	0	NONE	NONE
NATLIMR	8	real	Yes	0	NONE	NONE
NATLNG	8	real	Yes	0	NONE	NONE
NATLOG	8	real	Yes	0	NONE	NONE
NG	16	real	Yes	1	REGION	2
OG	16	real	Yes	1	REGION	2
POP	2752	real	Yes	3	AGE	86
					SEX	2
					REGION	2
POP5	576	real	Yes	3	AGE5	18
					SEX	2
					REGION	2
SEXRATIO	8	real	No	0	NONE	NONE
TFR	16	real	Yes	1	REGION	2
TOTALPOP	8	real	Yes	0	NONE	NONE
TOTPOP	16	real	Yes	1	REGION	2

Appendix to Chapter 3

Label Format Definitions

Format: AGE5

Description: Labels for five-year age groups

	Lower Bound	Upper Bound	Size of Interval (Inclusive)	or	Interval Label
1	0	84	5		
2	85	100			85+

Format: AGE611

Description: Labels for pop. 6 to 11 yrs.

	Lower Bound	Upper Bound	Size of Interval (Inclusive)	or	Interval Label
1	0	5			0 - 5
2	6	11			6 - 11
3	12	100			12+

Format: AGEGP

Description: Age group

	Lower Bound	Upper Bound	Size of Interval (Inclusive)	or	Interval Label
1	0	6			0-6
2	7	7			7
3	8	8			8
4	9	9			9
5	10	10			10
6	11	11			11
7	12	12			12
8	13	13			13
9	14	14			14
10	15	15			15
11	16	16			16
12	17	17			17
13	18	100			18-

Format: AGGPOP

Description: Labels for pop. aggreg. by 5-yr. groups

	Lower Bound	Upper Bound	Size of Interval (Inclusive)	or	Interval Label
1	0	0			0 - 4
2	1	1			5 - 9
3	2	2			10 - 14
4	3	3			15 - 19
5	4	4			20 - 24
6	5	5			25 - 29
7	6	6			30 - 34
8	7	7			35 - 39
9	8	8			40 - 44
10	9	9			45 - 49
11	10	10			50 - 54
12	11	11			55 - 59
13	12	12			60 - 64
14	13	13			65 - 69
15	14	14			70 - 74
16	15	15			75 +

Format: ASFR

Description: Label of age-specific fertility rates

	Lower Bound	Upper Bound	Size of Interval (Inclusive)	or	Interval Label
1	1	1			15-19
2	2	2			20-24
3	3	3			25-29
4	4	4			30-34
5	5	5			35-39
6	6	6			40-44
7	7	7			45-49

Format: BATCHQUE

Description: Questions for Batch run.

	Lower Bound	Upper Bound	Size of Interval (Inclusive)	or	Interval Label
1	1	1			ESTIMATE Method (1=Std, 2=Schief)
2	2	2			Default Grade 6 Rep. Rte MF
3	3	3			Default Grade 6 Rep. Rte Male
4	4	4			Default Grade 6 Rep. Rte Female

Format: BATCHQUE (continued)

5	5	5	Drop out proportion Male, Female
6	6	6	Drop out proportion Male
7	7	7	Drop out proportion Female
8	8	8	Last age disagg.? (1=no, 2=yes)
9	9	9	Aggregated Age
10	10	10	Disagg. by Age (1=disagg.,2=agg,)
11	11	11	Age Spec.Enrol.Rte? (1=yes, 2=no)
12	12	12	PROJRATE Meth. (1=Varying, 2=Fix)
13	13	13	Varying Meth(1=log, 2=lin, 3=tar)
14	14	14	Target year for Target Rates
15	15	15	Fixed Rates val. (1=mean, 2=last)
16	16	16	COHORT Meth (1=fixed, 2=varying)
17	17	17	Target Year for COST Module

Format: COEF_TYP

Description: Label for columns of res. coef. table

	Lower Bound	Upper Bound	Size of Interval (Inclusive)	or	Interval Label
1	1	1			Stud. per, base
2	2	2			Stud. per, trgt
3	3	3			Shifts, base
4	4	4			Shifts, target
5	5	5			Minimum level
6	6	6			Recurr. or cap.
7	7	7			Prop.

Format: COST_TYP

Description: Label for cost grades

	Lower Bound	Upper Bound	Size of Interval (Inclusive)	or	Interval Label
1	1	1			Type 1
2	2	2			Type 2
3	3	3			Type 3
4	4	4			Type 4
5	5	5			Type 5

Format: ENAGE

Description: Labels for enrolment age groups

	Lower Bound	Upper Bound	Size of Interval (Inclusive)	or	Interval Label
1	5	5			5 yr. olds
2	6	6			6 yr. olds
3	7	7			7 yr. olds
4	8	8			8 yr. olds
5	9	9			9 yr. olds
6	10	10			10 yr. olds
7	11	11			11 yr. olds
8	12	12			12 yr. olds
9	13	13			13 yr. olds
10	14	14			14 yr. olds
11	15	15			15 yr. olds
12	16	16			16 yr. olds
13	17	17			17 yr. olds
14	18	18			18 yr. olds
15	19	19			19 yr. olds
16	20	20			20 yr. olds
17	21	21			21 yr. olds
18	22	22			22 yr. olds
19	23	23			23 yr. olds

Format: FERTAGES

Description: Labels for range of fertile ages

	Lower Bound	Upper Bound	Size of Interval (Inclusive)	or	Interval Label
1	1	1			Min. fert. age
2	2	2			Max. fert. age

Format: FERTOPT

Description: Labels for fertility table options

	Lower Bound	Upper Bound	Size of Interval (Inclusive)	or	Interval Label
1	0	0			User ASFR
2	1	1			Sub-Saharan
3	2	2			Arab
4	3	3			Asian
5	4	4			User Pattern

Format: GRADE

Description: Labels for grade names

	Lower Bound	Upper Bound	Size of Interval (Inclusive)	or	Interval Label
1	1	1			First
2	2	2			Second
3	3	3			Third
4	4	4			Fourth
5	5	5			Fifth
6	6	6			Sixth
7	7	7			Seventh
8	8	8			Eighth
9	9	9			Ninth
10	10	10			10
11	11	11			11
12	12	12			12

Format: INT5OPT

Description: 1=INT5VAR structure, 2=INTVAR structure

	Lower Bound	Upper Bound	Size of Interval (Inclusive)	or	Interval Label
1	1	1			INT5VAR levels
2	2	2			INTVAR levels

Format: LTFILE

Description: Labels for life table files

	Lower Bound	Upper Bound	Size of Interval (Inclusive)	or	Interval Label
1	1	1			C-D North
2	2	2			C-D South
3	3	3			C-D West
4	4	4			C-D East
5	5	5			UN General
6	6	6			UN Lat. America
7	7	7			UN Far East
8	8	8			UN So. Asia
9	9	9			User-supplied

Format: LTROWS

Description: Labels for life-table rows

	Lower Bound	Upper Bound	Size of Interval (Inclusive)	or	Interval Label
1	1	1			Life Expectancy
2	2	2			0-4
3	3	3			5-9
4	4	4			10-14
5	5	5			15-19
6	6	6			20-24
7	7	7			25-29
8	8	8			30-34
9	9	9			35-39
10	10	10			40-44
11	11	11			45-49
12	12	12			50-54
13	13	13			55-59
14	14	14			60-64
15	15	15			65-69
16	16	16			70-74
17	17	17			75-79
18	18	18			80-84
19	19	19			85+

Format: MFCOLS

Description: Labels for model fertility age groups

	Lower Bound	Upper Bound	Size of Interval (Inclusive)	or	Interval Label
1	1	1			TFR
2	2	2			15-19
3	3	3			20-24
4	4	4			25-29
5	5	5			30-34
6	6	6			35-39
7	7	7			40-44
8	8	8			45-49

Format: MFROWS

Description: Labels for fert. levels in fert. tables

	Lower Bound	Upper Bound	Size of Interval (Inclusive)	or	Interval Label
1	1	1			TFR 1
2	2	2			TFR 2
3	3	3			TFR 3
4	4	4			TFR 4
5	5	5			TFR 5
6	6	6			TFR 6

Format: MODFERT

Description: Labels for fertility table regions

	Lower Bound	Upper Bound	Size of Interval (Inclusive)	or	Interval Label
1	1	1			Sub-saharan
2	2	2			Arab
3	3	3			Asian
4	4	4			User-supplied

Format: QUAL_TYP

Description: Label for cost grade distribution

	Lower Bound	Upper Bound	Size of Interval (Inclusive)	or	Interval Label
1	1	1			Type 1, base
2	2	2			Type 1, target
3	3	3			Type 2, base
4	4	4			Type 2, target
5	5	5			Type 3, base
6	6	6			Type 3, target
7	7	7			Type 4, base
8	8	8			Type 4, target
9	9	9			TYpe 5, base
10	10	10			Type 5, target

Format: R

Description: label for popfile

	Lower Bound	Upper Bound	Size of Interval (Inclusive)	or	Interval Label
1	1	1			ext (1) int (2)
2	2	2			filename

Format: REGION

Description: Labels for regions

	Lower Bound	Upper Bound	Size of Interval (Inclusive)	or	Interval Label
1	1	1			Total
2	2	2			Rural

Format: RES_TYPE

Description: Label for type of resource

	Lower Bound	Upper Bound	Size of Interval (Inclusive)	or	Interval Label
1	1	1			Teachers
2	2	2			Administrator
3	3	3			Other admin.
4	4	4			Learning res.
5	5	5			Classrooms
6	6	6			Laboratories
7	7	7			Land
8	8	8			Furniture

Format: SEX

Description: Labels for sexes

	Lower Bound	Upper Bound	Size of Interval (Inclusive)	or	Interval Label
1	1	1			Both sexes

CHAPTER 4. TECHNICAL DESCRIPTION OF THE MODEL METHODOLOGY

The Grade-transition Enrollment Projection Method

Of the many available methods for projecting enrollment, the grade-transition method, while perhaps not the most sophisticated, is a good compromise for developing countries since it does not require as much data as the more difficult methods, yet is useful for policy analysis.

The method, at the elementary school level, is straightforward, and consists of two basic equations. The first equation applies to grade 1 only:

$$E_{1,t+1} = r_{1,t}E_{1,t} + n_{t+1}POP_{entage,t+1}$$

Here $E_{1,t}$ is enrollment in grade 1 at time t, $r_{1,t}$ is the repetition rate in grade 1 at time t, n_{t+1} is the proportion of the population of age entage ("entrance age") that enrolls in grade 1 for the first time at time t+1, and $POP_{entage,t+1}$ is the population of age entage at time t+1. All these rates from now on will be collectively referred to as transition rates or paramaters. All these variables can be disaggregated by sex within the STEP system, but the sex subscript is left out of the technical description to simplify the notation.

This equation means that the number of students in grade 1, in any given year t+1, comes from two sources: a proportion $r_{1,t}$ of last year's enrollment in grade 1, $E_{1,t}$, which repeats the grade, and a proportion n_{t+1} of the population $POP_{entage,t+1}$ that enters school for the first time at age entage, in year t+1.

For grades 2-6, the basic equation is:

$$E_{g,t+1} = r_{g,t}E_{g,t} + (1-r_{g-1,t} - d_{g-1,t})E_{g-1,t}$$

where the symbols are analogous to those above, and $d_{g,t}$ is the dropout rate from grade g at time t.

That is, enrollment in year t+1 in any grade g comes from two sources: a proportion $r_{g,t}$ of last year's enrollment $E_{g,t}$, which repeats the same course, and a proportion $(1-r_{g-1,t} - d_{g-1,t})$ of last year's enrollment in the previous grade, $E_{g-t,t}$, that is promoted forward. Note that $(1-r_{g-1,t} - d_{g-1,t})$ is the same thing as the promotion rate.

This methodology can be used with either fixed or varying transition parameters. In either case, it is necessary to project the parameters themselves **before projecting enrollment.**

Projections with Fixed Parameters

Projecting with fixed parameters simply means that $r_{g,t} = r_{g,t-1}$ and $d_{g,t} = d_{g,t-1}$ for all grades g, and for all periods t. When projecting with fixed parameters, two intuitively appealing ways to fix the parameters are:

- at the historical mean, and
- at the last historical level

The computer program allows the user to choose between these. Simple as it may seem, projecting with fixed parameters is often the more correct option (see below).

Projections with Variable Parameters

In projecting with variable parameters, there are several common options:

- projecting with a linear trend
- projecting with a logistic trend
- projecting according to optimistic, medium, and pessimistic targets

Linear Trend. If one uses a linear trend to project the transition parameters, the equation is (using the repetition rate for any grade g as an example, but the method applies to any rate and any grade):

$$r_{g,t} = \begin{cases} a+bt & \text{if} \quad 0 \leq a+bt \leq 1 \\ 0 & \text{if} \quad a+bt < 0 \\ 1 & \text{if} \quad a+bt > 1 \end{cases}$$

where t is the time period, and the parameters a and b can be estimated using simple linear regression. Of course, if the statistical significance of these parameters, especially b, is questionable (one determines this by at least looking at its t-statistic - ideally the Durbin-Watson statistic should also be checked if available) it would be preferable to use fixed rates.

Logistic Curve Fitting. The problem with linear projections is a familiar one: there is nothing inherent in the method itself to prevent the projected rate from being greater than 1 or less than 0. Thus, it is necessary to impose those (or any other) bounds on it, as we have done above. Since this seems somewhat artificial (specially if the b coefficient is large and the rate is approaching either 1 or 0 fast), fitting a logistic curve, which automatically keeps the rate between 0 and 1, without restraining it artificially as it approaches those bounds, is usually a preferred method. Logistic curves are commonly used to model phenomena that are naturally bound from above and below, like the spread of disease or technology: the proportion of a population with a certain disease or using a certain technology cannot ever be below 0 or above 1. Furthermore, a logistic curve is somewhat "S"-shaped ("sigmoid"): at first, as it departs from the lower (or upper, if negatively sloped) bound, the slope is gentle, then it gets steeper, and as it approaches its upper (or lower, if negatively sloped) bound, the slope gets

gentle again. This seems like a good approximation to the time trend of a variable like a repetition rate: at first, it may be hard to reduce it because of lack of experience and ingrained habits. But with success comes learning and more success, so it gets easier. At the end, it gets difficult again, because the few remaining repeaters are "hard-core". The projecting equation would be:

$$r_{g,t} = \frac{1}{1+e^{a+bt}}$$

where e refers to the mathematical constant e. This implies that in estimating the parameters a and b, simple linear regression can be used, but only after first transforming the rate $r_{g,t}$ as follows:

$$\tilde{r}_{g,t} = \ln\left(\frac{1-r_{g,t}}{r_{g,t}}\right)$$

It is this rate which is then regressed on t to obtain a and b. The same cautions about the statistical significance of a, and specially b, apply here.

Target Setting Method. Since these two methods are somewhat mechanical, in that they simply extrapolate current trends, a method based on target-setting is often used. In this case, and assuming a linear approach to the target, the projection equation is

$$r_{g,t} = r_{g,t-1} + \frac{\bar{r}_{g,tt} - r_{g,lht}}{tt-lht}$$

where the bar in $r_{g,tt}$ indicates that this is a rate fixed by the user of the user method, tt refers to the target period, and lht refers to the last historical period. With this method, the user has to determine both the target rate and the target year. If the statistical projection equations discussed above are untrustworthy, because of the low significance of the slope coefficients b, then the target-setting method might be the more forthright. The method can be used to set optimistic, pessimistic, and medium targets.

Efficiency Indicators Based on Cohort Analysis

Basic Concept

The internal efficiency measures of a school system which are described in the next section are derived from calculating the school history of an artificial cohort of 1000 entrants to the school system and tracking them as they progress (or drop out, or remain) in the school system. The method consists of starting with a matrix of enrollment by grade and sex with 1000

students in grade 1 and nobody in grades 2 - 6. Here we assume a six year primary cycle for exposition purposes. The grade transition model is then repeatedly applied to this enrollment matrix until there is nobody left in the school system, i.e. the initial 1000 children have either dropped out or have graduated. As the grade transition model is iteratively applied, the number of students remaining in the system is tracked as well as the number of iterations. The number of iterations is conceptually equivalent to the amount of time the students remaining in the system have spent in school. Hence, if 100 of the 1000 entrants graduate "on time" after six years (i.e. they did not repeat), then they would be counted as graduates after six iterations of the method. Those members of the cohort who repeated a grade will obviously remain in the system for longer than six years and it will require more than six iterations for them to graduate, if at all.

In doing these calculations the number of student/years spent in the system for various categories of students is also computed. Thus the 100 "on time" graduates spent 600 student/years in the system. If 100 more students took seven years to graduate and no one else graduates, then the second group spent 700 years in the system and the total of 200 graduates spent 1300 student/years or an average of 6.5 years in the system per graduate. Similar calculations can be applied to students who leave the system without graduating.

Note that the method depends only on the drop out and repeater rates. For this reason the assumptions underlying the future transition rates during the future school history of the cohort are important. There are two basic methods which can be used here.

In the first method, all transition rates are assumed constant and equal to those during the entry year. In other words, the method projects the future history of a cohort entering school in 1986 according to the transition rates in 1986, and those rates are assumed to remain at 1986 levels.

In the second method, the transition rates change according to the actual "year" in the future projection of the cohort of entrants. In other words a cohort of 1000 entrants in 1986 would be subject to the transition rates in 1986 during the first iteration or "year", to the 1987 rates during the second iteration, to the 1988 rates during the third iteration, etc. If drop out rates are expected to fall over time, the second method would produce completion rates which are consistently higher and growing faster than those produced by the first approach.

Efficiency and Other Educational Indicators

Completion or Graduation Rate. Perhaps the most used, and certainly the simplest measure, this is the number of graduates per 1000 entrants.

"On Time" Graduates. This is the number of graduates per 1000 entrants who complete the prescribed cycle within the prescribed number of years.

Student/Years per Graduate. This is average amount of time spent in a school system per graduate. There are two methods for calculating this. In the first method, the number of student/years spent in the system by all graduates is divided by the number of graduates. In the second method, the number of student/years spent both by graduates and by drop-outs is divided by the number of graduates. In a six year system with repetition and drop out rates equal to zero these would be equal to six.

Input-output Ratio. A problem with the above measure is that is does not take into account the length of a cycle. A six year school system will require ideally six student/years per graduate which, from an efficiency standpoint should be no less efficient than an eight year system requiring ideally eight years. The input-output ratio standardizes for the length of the sysle by dividing the calculation for student/years per graduate by the length of the cycle. The ratio is ideally equal to one in any system.

Years Spent in School. This is the average number of years spent in school by all of the 1000 entrants before leaving the system. It is the average of the student/years per graduate and student/years per drop out, weighted respectively by the number of graduates and the number of drop outs. Note that the number of drop outs and the number of graduates must sum to 1000.

Indicators Based on Enrollment Data

Gross Enrollment Ratio This is the ratio of enrollment in primary school, in all age groups, divided by the population in the "legal schooling" age groups, usually 6-14 or 7-15, or thereabouts. Since the ratio is of actual or projected enrollment to actual or projected population, it has to be derived either from historical data or from an enrollment projection. It is not based on the analysis of an idealized cohort.

Net Enrollment Ratio This is a similar concept to the one above, except that the numerator is not all students, but only those in the same age groups as the population. Thus, when compared with the Gross enrollment Ratio, it is an indicator of the proportion of "overage" students in the system. It is important to note that this "overage" might be only in a legal sense: if a perfectly efficient system, with little repetition or dropouts, was unable to absorb the population right at the legal starting age, then the difference between the Net enrollment Ratio and the Gross enrollment Ratio would simply be a reflection of this fact, rather than a measure of inefficiency or wastage. In this sense, it differs from measures of "overage" derived from cohort analysis, such as the number of graduates not on time.

Estimation of Transition Rates

Standard Method

The simplest standard method of estimating the repetition and dropout rates involves using the official data on numbers of repeaters and enrollment to get the repetition rates, and solving the grade-transition equations for the dropout rates, using historical enrollment data and the previously estimated repetition rates.

The repetition rates are estimated as follows:

$$r_{g,t} = \frac{R_{g,t+1}}{E_{g,t}}$$

where $R_{g,t}$ is the number of repeaters in grade g, period t. Evidently, what we have here is just the ratio of repeaters to enrollment in the same grade the previous year. From this, we can go on to solve the grade-transition equations.

As we saw above, the standard equation of the grade-transition method for grades higher than 1 is:

$$E_{g,t+1} = r_{g,t}E_{1,t} + (1-r_{g-1,t} - d_{g-1,t})E_{g-1,t}$$

Solving $d_{g-1,t}$ for, we get

$$d_{g-1,t} = 1 - \frac{(E_{g,t+1} - r_{g,t}E_{g,t})}{E_{g-1,t}} - r_{g-1,t}$$

which is just 1 minus the promotion rate minus the repetition rate. Using this method we can solve for all the repetition and dropout rates. Note that the method implies that it will be easier to make the calculations if all repetition rates are calculated before the dropout rates, since to calculate the dropout rate for grade g we need to know the repetition rate not only for grade g, but also for grade g+1. Also note that the dropout rate implied in this method is the beginning-of-year to beginning-of-next-year dropout rate, i.e., it includes all those who abandon the school system between grades. Finally, it includes all those who die. This makes it unnecessary to refer to a death rate in making the enrollment projections.

The entrance (or first grade enrollment) rate, n_t, is calculated using the estimate repeater rate and the population and enrollment data:

$$n_{t+1} = \frac{E_{1,t+1} - r_{1,t}E_{1,t}}{POP_{entage,t+1}}$$

Schiefelbein Method - Description

The motivation for the Schiefelbein[2] method is the suspicion, confirmed in some cases, that the numbers of repeaters in most developing countries are under-reported. The reasons given for this are many:

- repetition is a social stigma, so children, when asked by a new teacher whether they are repeaters, are not likely to be 100% truthful
- some children who appear as new in a classroom are actually repeaters who have migrated from another region or school
- children may repeat a course even if they have not officially failed the previous one

It is easy to see from the equations for the standard method that, under these circumstances, the repetition rates will tend to be under-estimated, and the dropout rates consequently over-estimated.

The basic idea behind the Schiefelbein method is that there has to be a certain mathematical structure that determines the differences between two consecutive tables of enrollment by age and grade. This underlying mathematical structure must have age-specific repetition and dropout rates as its parameters. From the data on enrollment by age and grade for two consecutive years it may not be possible to uniquely identify a set of age-specific repetition and dropout rates, but, with certain unrestrictive assumptions, one can at least obtain a range within which the overall (i.e., not age-specific) rates MUST lie, given the age-grade enrollment data and the underlying mathematical structure.

The method consists of the following. We assume a primary school system consisting of six grades, an entrance age of 7, and a maximum school age of 23. These assumptions are for ease of exposition, but the method is easily generalizable. Also, the method can be used to calculate rates disaggregated by sex, if the enrollment data are available by sex. In the discussion below we abstract from this to keep things a little simpler.

We start with age-grade enrollment data for two consecutive years. Thus, we have $E_{g,a,t}$ and $E_{g,a,t-1}$, where g stands for grade, a stands for age, and t stands for time.

The calculations then proceed grade by grade, for each pair of consecutive ages. For any age group a, therefore, we first make the calculations for grade 6. We start by calculating the number of repeaters of age a, in grade 6 at time t.

$$R_{6,a,t} = r_6 E_{6,a-1,t-1}$$

where r_6 is an assumed value. From here we can then go on to calculate the following. The total number of students promoted from grade 5 to grade 6 has to be:

---

2. See Schiefelbein, Ernesto, "Statistical Report on Repetition in Latin America," report prepared for UNESCO, Division of Statistics on Education, Office of Statistics, Paris, 1980.

$$P_{6,a,t} = \begin{cases} E_{6,a,t} - R_{6,a,t} & \text{if} \quad E_{5,a-1,t-1} + R_{6,a,t} \geq E_{6,a,t} \\ E_{5,a-1,t-1} & \text{otherwise} \end{cases}$$

In words: if the sum of enrollment in grade 5 last year, plus those repeating grade 6 this year, is greater than enrollment in grade 6 this year, then we have to put a cap on how many students could have been promoted from grade 5 to grade 6. This cap is enrollment in grade 6 this year, minus repeaters in grade 6 this year. But if the sum of enrollment in grade 5 last year, plus those repeating grade 6 this year, is less than enrollment in grade 6 this year, then we have to promote everyone who was in grade 5 last year, unless we are ready to admit that there were a lot of students new to grade 6, which would not be logical.

The number of new students in grade 6[3] will then be:

$$N_{6,a,t} = \begin{cases} E_{6,a,t} - P_{6,a,t} & \text{if} \quad E_{5,a-1,t-1} + R_{6,a,t} \leq E_{6,a,t} \\ 0 & \text{otherwise} \end{cases}$$

which follows from the determination of the number of promoted: if there were not enough people in grade 5 last year, such that the sum of enrollment in grade 5 last year, plus repeaters in grade 6 this year is less than enrollment in grade 6 this year, then there must have been some new students in grade 6 to make up the difference.

Having done the calculations for grade 6, we can then proceed with the earlier grades, still working with the same pair of consecutive age groups. The number of non-promoted students, that is, students who will either repeat any given grade g or drop out, will be:

$$NP_{g,a,t} = E_{g,a-1,t-1} - P_{g+1,a,t}$$

That is, the number of non-promoted from grade g is equal to the enrollment in that grade, minus those who were promoted to a higher grade. (Note that this implies that the calculation has to proceed from the higher grades to the lower ones.)

---

3. Note: for national-level projections the only new students in grade 6 can be international migrants. In regional projections, they may be inter-regional migrants.

The number of repeaters is calculated according to:

$$R_{g,a,t} = \begin{cases} (1\text{-dropprop})NP_{g,a,t} & \text{if} \quad NP_{g,a,t} \leq E_{g,a,t} \\ E_{g,a,t} & \text{otherwise} \end{cases}$$

where the proportion dropprop determines the proportion of non-promoted students who drop out. The number of repeaters in grade g is equal to a proportion of the non-promoted if the pool of non-promoted is less than enrollment, but if there are more non-promoted students from the previous grade than enrollment in grade g, then the maximum number of repeaters is set equal to enrollment. Note that the proportion dropprop plays a role, but only under certain conditions (the first condition in the equation above). The smaller the number of cases where those conditions are met, the less dependent on the particular assumed value of dropprop will be the results.

The number of students promoted to grade g will be:

$$P_{g,a,t} = \begin{cases} E_{g,a,t} - R_{g,a,t} & \text{if} \quad E_{g-1,a-1,t-1} + R_{g,a,t} \geq E_{g,a,t} \\ E_{g-1,a-1,t-1} & \text{otherwise} \end{cases}$$

The number of new students in grade g will be:

$$N_{g,a,t} = \begin{cases} E_{g,a,t} - P_{g,a,t} - R_{g,a,t} & \text{if} \quad E_{g-1,a-1,t-1} + R_{g,a,t} \leq E_{g,a,t} \\ 0 & \text{otherwise} \end{cases}$$

That is, if the pool of enrollment in the previous grade last year, plus repeaters of grade g, is larger than enrollment in grade g this year, then the number promoted has to be limited, and the limit is set to equal enrollment in grade g minus repeaters of grade g. But if the the number of repeaters of grade g plus enrollment in grade g-1 last year is smaller than enrollment in grade g this year, then we set the number promoted to equal the number enroled in g-1 last year.

And, finally, the number of dropouts can be calculated as:

$$D_{g,a,t} = NP_{g,a,t} - R_{g,a,t}$$

or the difference between the total that does not go on to grade g+1 and those who repeat grade g.

The equations for grade 1 are slightly different only because there is no previous grade. We list them for completeness only.

$$NP_{1,a,t} = E_{1,a-1,t-1} - P_{2,a,t}$$

$$R_{1,a,t} = \begin{cases} (1\text{-dropprop})NP_{1,a,t} & \text{if} \quad NP_{1,a,t} \leq E_{1,a,t} \\ E_{1,a,t} & \text{otherwise} \end{cases}$$

$$P_{1,a,t} = 0$$

(since there cannot be anyone promoted to grade 1)

$$N_{1,a,t} = \begin{cases} E_{1,a,t} - R_{1,a,t} & \text{if} \quad R_{1,a,t} \leq E_{1,a,t} \\ 0 & \text{otherwise} \end{cases}$$

$$D_{1,a,t} = NP_{1,a,t} - R_{1,a,t}$$

After having calculated all these flows, we finally get the transition paramaters as follows:

$$r_{g,t} = \frac{\sum_{a=5}^{23} R_{g,a,t+1}}{\sum_{a=5}^{23} E_{g,a-1,t}}$$

$$d_{g,t} = \frac{\sum_{a=5}^{23} D_{g,a,t+1}}{\sum_{a=5}^{23} E_{g,a-1,t}}$$

$$n_t = \frac{\sum_{a=5}^{23} N_{1,a,t}}{POP_{entage,t}}$$

Thus, the grade-specific transition rates are in fact weighted averages of the age-and-grade-specific transition rates. The presumption is that this averaging process will smooth out some of the irregularities inherent to the methodology, which arise because the data are not enough to uniquely identify all the age-specific transition rates.

Schiefelbein Method - Example of Calculations

Since the method is somewhat involved, a numerical example might help to understand the calculations required.

The following example uses data from Guatemala, from the years 1982 and 1983, for the age groups 11 (in 1982) and 12 (in 1983).

Grade	$E_{11,1982}$	P	NP	R	D	N	$E_{12,1983}$
First	17737		6477	6348	130	8299	14647
Second	21605	11260	6202	6078	124	0	17338
Third	24863	15403	5613	5500	112	0	20903
Fourth	23167	19250	2224	2180	44	0	21430
Fifth	19964	20493	457	448	9	0	21391
Sixth	6044	19507		302	0	0	19809

We start with the number of repeaters in grade 6, 302, which is 6044 multiplied by the assumed grade 6 repetition rate (0.05 here). This appears in the R column. Since we have 302 students repeating grade 6, and since 19964+302>19809, we cannot promote everyone that was in grade 5 in 1982. We therefore only promote 19809-302=19507. They appear in the P column. If grade 5 enrollment in 1982 had been, say, 19000, then, since 19000+302<19809, we would have promoted all 19000, and the difference between 19809 and 19302, which is 506, would have had to be new students. By promoting everyone when we have a situation like 19000+302<19809 we minimize the need to suppose there are new students entering grade 6.

Since 19507 of the 19964 students in grade 5 in 1982 were promoted, 457 (=19964-19507) must not have been promoted, so they appear in the NP column. These 457 students are split into repeaters and dropouts according to the assumed ratio dropprop, mentioned above, but only if the number of non-promoted is smaller then enrollment in grade 6 in 1983. Since 457 is indeed smaller than 21391, then the ratio is used in this case. Had we had, say, 23000 non-promoted (from grade 5) students, then only 21391 could have been repeaters, and the rest would have been dropouts. Now we can calculate the number promoted to grade 5. Since the enrollment pool in grade 4 in 1982, plus the number repeating grade 5 in 1983, is sufficient to supply the 22229 in grade 5 in 1983 (23167+448>21391), we need to cap the number that were promoted. Only 20493 (=21391-448) could have been promoted. The number of new students is 0, for grade 5 in 1983, because enrollment in grade 4 in 1982 (23167), plus those repeating grade 5 (448), was sufficient to supply the 21391 that were enroled in grade 5 in 1983 (since 23167+448>21391).

The process continues upward until we complete this table. Similar tables have to be completed for all consecutive age groups. This gives us total numbers of repeaters and dropouts by age. We aggregate these, and aggregate the enrollment as well, divide the repeaters and dropouts by enrollment, and get the repetition and dropout rates.

Projections of Resource Requirements and Costs

The COST module uses common formulas for projecting resource requirements and costs. This makes it possible to add cost categories for specific countries, as long as the methodology used to calculate costs as outlined here is acceptable.

First, total resource use by type of resource is calculated as follows:

$$RES_USE_{j,t} = TOTENROL_t \ / \ (RES_COEF_{j,t} * SHIFTS_{j,t})$$

for recurrent costs except the "other" resources category, and where $RES_USE_{j,t}$ is total requirement of resource j at time t, $RES_COEF_{j,t}$ is the number of students per unit of resource j at time t, and $SHIFTS_{j,t}$ is the number of times per day that resource j is used at time t. Physical requirements for the "other" resources category are not calculated since because the concept is irrelevant. As most variables in the cost calculations, all of the variables in this formula can be disaggregated by sex, but the sex subscript is left out to simplify the notation. Both RES_COEF and SHIFTS for every time period between the base year and the target year are interpolated by the module from the base year and target year values given by the user in the Host variable RES_COEF. For every year after the target year these variables remain at the target year level. For capital costs (i.e., investment) the formula is

$$RES_USE_{j,t} = (TOTENROL_{t+1} - TOTENROL_t) \ / \ (RES_COEF_{j,t} * SHIFTS_{j,t})$$

Note that according to this expression investment requirements might become negative if enrollment declines. This can have several interpretations. It is not likely that there would be significant disinvestment in the short run in any real situation. Thus, the simplest interpretation is to take any negative numbers here as a measure of slack in the use of the existing resource base.

Second, physical resource use or requirements need to be converted into monetary costs. This is done in several steps.

Unit costs by type of resource and cost level (or salary grade) are projected as

$$UNITCOST_{j,k,t} = UNITCOST_{j,k,t-1} * (1 + COSTINCR_{j,k})$$

where COSTINCR is the yearly cost increase rate for resource j of cost grade k.

Average unit cost calculations for each resource j, weighted by the distribution across cost categories, must then be made:

$$AVUNITCOST_{j,t} = \sum_{k} QUAL_{j,k,t} * UNITCOST_{j,k,t}$$

where $QUAL_{j,k,t}$ is the distribution of resource j across cost or salary grades k at time t. This is interpolated by the module from the base year and target year distributions given by the user in Host variable QUALDIST.

Once the average unit costs are known, the cost of the resources can be calculated according to the equation

$$RES_COST_{j,t} = RES_USE_{j,t} * AVUNITCOST_{j,t}$$

for all resources except the "other" resources category.
For the "other" category, costs are calculated as

$$RES_COST_{other,t} = (PROP/(1-PROP)) \sum_{j \neq \text{"other"}} RES_COST_{j,t}$$

where PROP is the ratio of "other" costs to total costs.

Finally, two per-student cost indicators are calculated: cost per capita (COST_CAP) and cost per graduate (COST_GRA). Cost per student is calculated according to the equation

$$COST_CAP_t = \sum_{j} RES_COST_{j,t} / TOTENROL_t$$

If resource costs for a given category of resources are negative they are excluded from the above sum. Cost per graduate is calculated according to the formula equation

$$COST_GRA_t = \sum_{r=0}^{5^+} (COST_CAP_{t+r} * SYGRAD2_{t+r})/6$$

where $SYGRAD2_t$ is the number of student-years it takes to produce one graduate at time t. Thus, if $COST_CAP_t$ is the cost/student-year at time t, and $SYGRAD2_t$ is the student-years/graduate at time t, then $COST_CAP_t*SYGRAD2_t$ is the cost per graduate at time t. But since each graduate starting at time t will be in the system for at least 6 years, we take the average of the costs per graduate for the 6 years starting with year t. Note that this implies that one can only produce data on the cost per graduate for 6 years less than there are data for the cost per student. Note also that in fact the average number of years it takes a graduate to get through the system is a variable greater than 6 and thus one would ideally take the average of a variable number of years. The disadvantage would be that then one would be able to produce even less data on cost per graduate. Using 6 gives a good approximation without losing too much data.

Projecting Population

POP uses the cohort-component method to project the base population forward. This method has been adapted to make annual, single-year age group population projections. The base-year population can be entered either in single-year age groups, if available, or in five-year age groups and then disaggregated using Beers multipliers. The survival ratios used are calculated from the disaggregated (again, using Beers coefficients) L values of a life table.

The basic projection equation is

$$POP_{a,s,r=1,t} = SURVIVORS_{a,s,r=1,t} + INTMIG_{a,s,r=2,t} - INTMIG_{a,s,r=1,t} + EXTMIG_{a,s,r=1,t}$$

for region 1, and

$$POP_{a,s,r=2,t} = SURVIVORS_{a,s,r=2,t} + INTMIG_{a,s,r=1,t} - INTMIG_{a,s,r=2,t} + EXTMIG_{a,s,r=2,t}$$

for region 2.

The components are calculated as follows.

First, the survivors are determined from last year's population. There are three relevant equations. For those newly born,

$$SURVIVORS_{a=0,s,r,t} = SURVIVALRATE_{a,s,r,t} * BIRTHS_{s,r,t},$$

then for the age groups between the first and the last,

$$SURVIVORS_{a,s,r,t} = SURVIVALRATE_{a-1,s,r,t-1} * POP_{a-1,s,r,t-1}$$

and for the last age group,

$$SURVIVORS_{a,s,r,t} = SURVIVALRATE_{a-1,s,r,t-1} * (POP_{a-1,s,r,t-1} + POP_{a,s,r,t-1})$$

The survival rates in all these equations are determined as the ratio between the Lx values for successive age groups in the life tables chosen by the user. But the Lx values needed here are for single-year age groups. Thus, the Lx values from the standard five-year tables (Coale-Demeny and U.N.) are first interpolated, using Beers coefficients, into single-year values. The justification for this is that the Lx values from the standard five-year life tables are conceptually equivalent to a stable population, and thus using Beers coefficients to interpolate them is equivalent to using Beers coefficients to interpolate population data grouped in five-year age groups into population data grouped into single-year age groups. However, whatever

objections can be made against using Beers coefficients for interpolating population values (especially in the early and late age groups) also hold against using them for interpolating the Lx values from the life tables. After interpolating using Beers coefficients to get single-year age group values, the POP module interpolates linearly to get Lx values appropriate for the life expectancies entered by the user under the variable LEXP.

In the first equation for calculating the number of survivors above, the variable BIRTHS is used. The number of births is calculated according to

$$BIRTHS_{male,r,t} = (SEXRATIO/(1 + SEXRATIO)) * \sum_{a=15}^{49} ASFR_{a,r,t} * POP_{a,female,r,t-1}$$

$$BIRTHS_{female,r,t} = (1/(1 + SEXRATIO)) * \sum_{a=15}^{49} ASFR_{a,r,t} * POP_{a,female,r,t-1}$$

where SEXRATIO is the ratio of boys to girls at birth, and ASFR is the age-specific fertility rate, disaggregated in for single-year age groups, disaggregated assuming births occur uniformly within each five-year age group in the fertile ages.

Internal migrants are calculated from a total migration rate given by the user by region only, and then disaggregated by age and sex within the program according to a model function. The equation for the model is

$$INTMIG_{a,s,r,t} = MIGRATE_{r,t}$$

$$* \; m_{a,s} * SURVIVORS_{a,s,r,t} / (\sum_{a} m_{a,s} * SURVIVORS_{a,s,r,t}))$$

$$* \sum_{a} \sum_{s} SURVIVORS_{a,s,r,t}$$

where $m_{a,s}$ is a function of age a and sex s. The function for $m_{a,s}$ is

$$m_{a,s} = a1_s * \exp(-alpha1_s * a)$$

$$+ a2_s * \exp(-alpha2_s * (a-mu2_s) - \exp(-lambda2_s * (a-mu2_s)))$$

$$+ c_s$$

where a is age, and a1, alpha1, a2, alpha2, mu2, lambda2, and c in turn are functions of sex:

	Male	Female
a1	0.01992	0.02209
alpha1	0.10390	0.10883
c	0.00263	0.00350
mu2	19.67385	19.88280
a2	0.06471	0.06935
alpha2	0.10618	0.13434
lambda2	0.37244	0.37870

Finally, for external migrants it is assumed that only total numbers are known to the user. The number of migrants by age, sex and region are calculated simply by assuming that the distribution by age, sex and region is the same as for the population:

$$EXTMIG_{a,s,r,t} = EXTMIG_t$$

$$* \; POP_{a,s,r,t-1} / \left(\sum_a \sum_s \sum_r POP_{a,s,r,t-1} \right)$$

This is evidently not the best solution, but there is no model function for external migration equivalent to the one used for internal migration.

Appendix to Chapter 4
Glossary of Variables and Symbols

The following table shows, for each variable, its symbol in this technical description, its name in the user's manual or in the computer program, and a brief description of its meaning. The variables are listed in the alphabetical order of their symbols.

Symbol	Name	Meaning
$COSTINCR_{j,k}$	COSTINCR	rate of cost increase for input j, cost or salary grade k
$d_{g,t}$	DROPRT	dropout rate grade g, time t
dropprop	DROPPROP	assumed share of dropouts in non-promoted students
$E_{g,a,t}$	ENROLAGE	enrollment in grade g, age a, at time t
$E_{g,t}$	ENROL	enrollment grade g, period t
endage	ENDAGE	oldest "legal" school age
entage	ENTAGE	age of entry into schooling
n_t	ENTRT	rate of entry from pop. to 1st grade
$POP_{a,t}$	POP	population of age a at time t
$r_{g,tt}$	TARREP	target repetition rate
$R_{g,t}$	NREPEAT	repeaters, grade g, time t
$QUAL_{j,k,t}$	QUALDIST	proportion of input j of quality level k at time t
$r_{g,t}$	REPEATRT	repeater rate grade g, time t
$RES_COEF_{j,t}$	RES_COEF	resource use and other coefficients for resource j at time t
tt	TARYEAR	target time period
$UNITCOST_{j,k,t}$	UNITCOST	cost per unit of input j, cost or salary grade k at time t
None	AGGENROL	enrollment in the aggregated age group
None	ASER	age-specific enrollment rate
None	COSTTYPE	cost categories and descriptions
None	GER	gross enrollment ratio
None	GRADRATE	number who graduate out of 1000 entrants
None	HISTORIC	indicator var.: historic=h, projected=p
None	INCLUDE	indicator var.: if include this year, include=y, if not, include=n.
None	IORATIO	input-output ratio: student years per graduate divided by the nominal school cycle length
None	ONTIME	number of "on-time" graduates, i.e., have not repeated
None	SCHLPOP	total (aggregated across ages) population of school age
None	SYGRAD	student/years per graduate not counting student/years "wasted" by drop-outs

Glossary of Variables and Symbols (continued)

None	SYGRAD2	student/years per graduate, counting student/years "wasted" by drop-outs
None	TARDROP	target dropout rate
None	TDROP	t-statistic for drop. rate regr. slope
None	TOTENROL	total (aggregated across grades) enrollment in primary school
None	TREP	t-statistic for rep. rate regr. slope
None	YRSSCHL	average no. of yrs. spent in school

The following variables are used only for intermediate calculations, and so the user need not be concerned about them. Within the program, some of these variables are simply various columns of the matrix "flowmatrix".

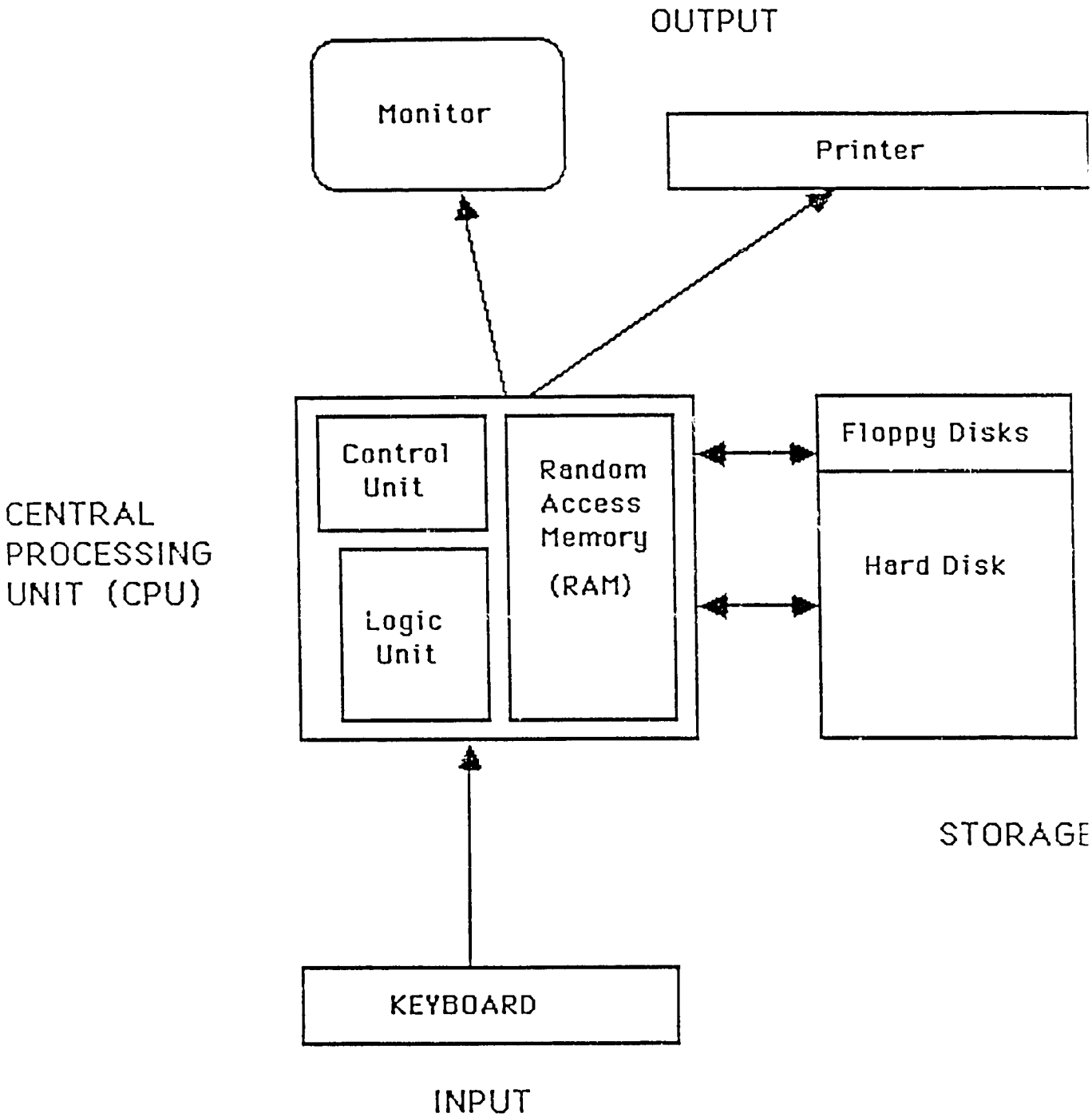
Symbol	Name	Meaning
a	various	regression intercepts
b	various	regression slopes
$D_{g,a,t}$	flowmatrix	dropouts from grade g, of age a, at time t
lht	lasthist	last historical period
$N_{g,a,t}$	flowmatrix	new students in grade g, of age a, at time t
$NP_{g,a,t}$	flowmatrix	students not promoted from grade g, of age a, at time t
$P_{g,a,t}$	flowmatrix	students promoted to grade g, of age a, at time t
$R_{g,a,t}$	flowmatrix	repeaters of grade g, of age a, at time t
t	t	time period

The following variables are used in the population projection database only.

Symbol	Name	Meaning
$ASFR_{a,rt}$	ASFR	age specific fertility rates for age a, region r, at time t
$ESTMIG_{a,s,r,t}$	EXTMIG	external migration in absolute numbers, for age a, sex s, region r, at time t
$MIGRATE_{r,t}$	MIGRATE	internal migration rate for region r at time t
$POP_{a,s,r,t}$	POP	population by single-year age groups, age a, sex s, region r, at time t
SEXRATIO	SEXRATIO	ration of boys to girls at birth
NONE	FEMINDEX	index denoting women in internal calculations
NONE	LEXP	life expectancy
NONE	LT	life table L values
NONE	MODFERT	model fertility schedules
NONE	MORTOPT	mortality options
NONE	TFR	total fertility rate

Section V: Microcomputer Information

DIAGRAM OF A MICROCOMPUTER SYSTEM



INTRODUCTION/REVIEW OF BASIC LOTUS 1-2-3 COMMANDS

*** Table of Contents ***

INTRODUCTION

(1) STARTING A LOTUS WORKING SESSION

(2) ENDING A LOTUS WORKING SESSION

(3) SAVING WORKSHEETS

(4) RETRIEVING FILED WORKSHEETS

(5) ERASING WORKSHEET FILES

(6) MOVING AROUND IN THE WORKSHEET

(7) DEFINITION OF RANGES

(8) CORRECTING ERRORS

(9) TYPES OF CELL ENTRIES

(10) ENTERING DATA

A. Entering Labels

B. Entering Values

C. Entering Formulas

(11) COPYING CELL CONTENTS TO OTHER CELLS

(12) ADJUSTING COLUMN-WIDTHS

(13) COLUMN AND ROW INSERTION

(14) PRINTING

(15) ARITHMETIC OPERATIONS

(16) ARITHMETIC FUNCTIONS

(17) EXAMPLES OF FORMULAS

INTRODUCTION

Lotus 1-2-3 comes on a set of floppy diskettes which can be used separately on a floppy disk based computer system, or they can all be loaded onto a hard disk system so that all of the programs are available to you at once. The various programs on these diskettes enable you to create worksheets and graphs and print them out. One floppy disk contains a tutorial called "A View of 1-2-3."

The instructions that follow assume that Lotus has been installed on your computers hard disk in a directory called LOTUS.

(1) STARTING A LOTUS WORKING SESSION

1 - Turn on the power to the system.

2 - Enter the date and time.

3 - Type **CD\Lotus** to change to the Lotus directory.

4 - At the prompt, C:\Lotus> , type **Lotus** then press the **ENTER** key to activate the Lotus system.

5 - Select **123** from the Lotus System Menu to activate the Lotus spreadsheet.

[Note: Selection of <u>any</u> command from <u>any</u> Lotus menu is made by moving the cursor key to the desired function and then pressing **ENTER**.]

+ A portion of the Lotus spreadsheet will appear on the screen. You are now ready to build a Lotus worksheet.

(2) ENDING A LOTUS WORKING SESSION

To end a working session, you must execute the QUIT command from the Lotus command menu. However, the QUIT command does not save your work. You must save work before quitting (see (2), below.) To end a session:

1 - Press the / key to activate the Main Commands Menu.

2 - Select **QUIT** from the Main Menu.

3 - Press the **ENTER** key.

+ The computer will respond with the following prompt with NO highlighted by the cursor:

<u>NO</u> YES
Do not end 1-2-3 session; Return to READY mode

4 - Select **YES** and then press the **ENTER** key to end the working session and return to the Lotus System Menu.

[Note: Selecting NO would return you to your worksheet. If you are not sure if saved your work or if you wish to continue working, select NO.]

5 - Select **EXIT** and then press the **ENTER** key.

+ The prompt, C:\Lotus> , will appear on the screen. You have ended the work session and finished with Lotus.

(3) SAVING WORKSHEETS

You should save your work on disk periodically as you go along. If the electrical power to your computer goes off, your worksheet will be lost unless you have it saved on a disk.

1 - Press the / key to activate the Main Commands Menu.

2 - Select **FILE** then press **ENTER** to activate the File Menu.

3 - Select **SAVE** from the File Menu then press **ENTER**.

+ In response, the computer will issue the following prompt:

Enter save file name:

4a - If this is the first time you are saving a worksheet, press the **ESC** key <u>two</u> times then type in the name you wish to call your worksheet and press the ENTER key.

[Note: the name you use must include the location to which you want to save the file. For example, to save a worksheet to a file called PUPILS on a disk in the A drive, you would type **A:PUPILS** then **ENTER**.]

4b - If you are saving an edited version of a worksheet which you have previously saved, Lotus will display the name of the worksheet after the prompt, for example:

Enter save file name: A: PUPILS

- If you want the worksheet to be saved under its' original name, press the **ENTER** key.

+ After pressing the ENTER key, the following prompt appears with **Cancel** highlighted:

Cancel Replace
Cancel Command --- **Leave existing file intact**

- To save the current worksheet using its' original name select **Replace** then **ENTER.**

+ Selecting Cancel returns you to the worksheet.

4c - If you want a previously saved worksheet to be saved under a new name:

- in response to the prompt, **Enter save file name:** press the **ESC** key three times then type in the name you wish to call your worksheet and press the **ENTER** key.

IMPORTANT: Always save a file before retrieving another. Otherwise, the original will be erased permanently by the retrieval process.

(4) RETRIEVING FILED WORKSHEETS

To retrieve a worksheet previously stored on your disk:

1 - Press the / key.

2 - Select **FILE** then press **Enter.**

3 - Select **RETRIEVE** then press **ENTER.**

+ In response, the computer will issue the following prompt:

Name of file to retrieve: C:\Lotus*.wk?

+ Below this prompt will be a list of worksheet files saved on the default drive with the first file highlighted.

4a - If you want to use a file saved on the default drive, either:

1 - Using the arrow keys select the file you want and then press **ENTER,** or

2 - Type the name of the file and then press **ENTER.**

4b - If you want to use a file saved on a disk other than the default disk:

- Press the **ESC** key three times, then type the name of the complete file name including the location (see Note, Section (3)-4a).

(5) ERASING WORKSHEET FILES

To erase a worksheet file from the screen (not to be confused with deleting it from the disk):

1- Press the / key.

2 - Select **FILE** then **ENTER.**

3 - Select **ERASE** then **ENTER.**

4 - Select **WORKSHEET** then **ENTER.**

+ The computer responds with the prompt:

Enter name of file to erase:

+ Below this prompt will be a list of worksheet files saved on the default drive with the first file highlighted.

5a - If you want to erase a file saved on the default drive, either:

1 - Using the arrow keys select the file you want and then press **ENTER**, or

2 - Type the name of the file and then press **ENTER.**

+ Lotus responds with the following prompt with **NO** highlighted.

No Yes
Cancel command --- Do not erase the file

- If you definitely want to erase the worksheet file, then press **Y** to confirm your decision. If you change your mind and decide to keep the file, press **N** to cancel the command.

5b - If you want to erase a file saved on a disk other than the default disk:

- Press the **ESC** key three times, then type the name of the complete file name including the location (see Note, Section (3)-4a).

(6) MOVING AROUND IN THE WORKSHEET

The Lotus 1-2-3 WORKSHEET is composed of **cells**. The location of a cell, also called a "cell address," is determined by its relation to the letters across the top of the worksheet and the numbers down the left side of the worksheet. The cell pointer is the highlighted rectangle in the worksheet. It indicates the location of the currently active cell. As you move the cell pointer around the worksheet, the cell address indicator located at the top left of the screen will change accordingly. There are many ways to change the position the cell pointer throughout the worksheet:

[NOTE: All these maneuvers describe the response of Lotus with the **Scroll Lock** key turned OFF. You may sometimes get some unexpected responses if it is on.]

1. Press the **ARROW** keys to move the cell pointer one space in any direction.

2. Press the **HOME** key to return the cell pointer to the coordinate position A1.

3. Press the **END** key to move the cell pointer to the next boundary of contents.

4. Press the **END key and then the HOME key** to move the cell pointer to the end or bottom right-hand corner of the **active area** of the worksheet. The active area refers to the area of the worksheet you have used.

5. Press the **GOTO** key (that stands for "GO TO"), which is the **F5 function key**, and type in the cell address (location or column/row designation) you wish the cell pointer to go to and press the ENTER key.

There are also ways to move throughout the worksheet by pages, or full screens at a time:

1. Press the **PgUp** key to move one full screen of the worksheet up.

2. Press the **PgDn** key to move one full screen of the worksheet down.

3. Press the **TAB** key to move one full screen of the worksheet to the right.

4. Hold down the **UPPERCASE** (or **SHIFT**) key and simultaneously press the **TAB** key to move one full screen of the worksheet to the left. In other words, the SHIFTED TAB acts in the reverse direction as the TAB key by itself.

(7) DEFINITION OF RANGES

One or more cells that form a rectangular block is called a **RANGE.** Many Lotus commands require the use of ranges to specify cells to be used in formulas, copied, erased, etc... When a RANGE specification is required Lotus displays a prompt that begins with **Enter range** and ends with a specification of coordinates. For example:

Enter range to format: A1..A1 or

Enter range to copy: B5..B5

A1..A1 and B5..B5 indicate that the range currently contains only one cell, A1 and B5, respectively. They also indicate that, in the first case, the cell pointer was located in cell A1 when the range prompt appeared. In the second case the cell pointer was located in cell B5.

In most cases, you will want to specify a range that contains more than one cell. For example assume you want to specify the range of all the cells within the coordinates B8 and D12. There are two ways to do this: by **typing-in** the cell coordinates or by using the **pointing method.**

<u>The Typing-In Method</u>

1 - In response to the range prompt, press the **ESC** key, then type the first range coordinate, for example **B8.**

2 - Type .. (two periods).

3 - Type **D12.**

4 - Press **ENTER** to complete the range specification.

<u>The Pointing Method</u>

1 - In response to the range prompt, press **ESC** once.

2 - Using the arrow keys move the cell pointer to the first cell in the range, cell **B8.**

3 - Press the . key (Period key).

4 - Using the arrow keys move the cell pointer over and down to the second location coordinate, cell **D12.**

+ The range will be highlighted on the screen.

5 - Press the **ENTER** key

+ The range specification is completed.

(8) CORRECTING ERRORS

Lotus 1-2-3 provides many ways to correct errors.

1. A command can be revoked by pressing the **ESC** key. Or a series of commands revoked by pressing the **ESC** as many times as necessary to move backwards trough the Lotus menus to the desired command.

2. Mistakes can be erased by using the **RANGE-ERASE** command.

   1 - Press the / key.

   2 - Select **RANGE** and then **ENTER.**

   3 - Select **ERASE** and then **ENTER.**

   + Lotus responds with the following prompt:

   Enter range to erase:

   4 - Specify the range to erase as in section (7), above.

   5 - Press **ENTER** to erase the specified range.

3. The **BACKSPACE** key and **DELETE** key can be used to correct errors before you press ENTER (i.e., before the errors are entered).

4. The contents of a cell can be completely replaced by typing over the contents or by moving other information into that cell.

5. The contents of cells can also be edited y activating the Edit Mode. To enter the **EDIT MODE:**

   1 - Position the cursor at the cell you want to edit.

   2 - Press the F2 key.

   + The contents of the selected cell will appear on the second line , top left corner of the screen.

   3 - Use the arrow keys to move the cursor through the cell contents.

   4 - Use the **BACKSPACE** and **DELETE** keys to erase unwanted entries.

   [NOTE: The Backspace key erases the character to the left of the cursor position. The Delete key erases the character located at the position of the cursor.]

5 - Additions to the cell are made by positioning the cursor in the desired position and typing.

6 - Press the **ENTER** key when editing is completed.

(9) TYPES OF CELL ENTRIES

There are two types of data which may be entered in the cells: **VALUES** and **LABELS**. Lotus 1-2-3 distinguishes between these on the basis of the type of data you enter in the cell. When entering data, the first key pressed determines which of the two types the entry will be.

The **VALUE** type of cell entry includes both numerals and formulas. Numbers may begin with any digit 0 to 9, a plus (+) or minus (-) sign, a period (.), or dollar sign ($). A **VALUE** can end with a percent sign (%), which means that a number is divided by 100. Only one decimal sign can be inserted. Commas cannot be included.

The **LABEL** type of cell entry is used for headings and explanations. **LABELS** may begin with any remaining character not defined as a VALUE.

(10) ENTERING DATA

A. Entering Labels

1 - Position the cell pointer in the desired cell using the arrow keys.

2 - Type the desired label and then press the **ENTER** key. A label can be anything you want it to be. Labels are primarily used as headings or descriptions.

+ The default alignment for labels is flush left or the label preceded by ', an apostrophe.

+ To align labels right, precede the entry with a ", a double quote.

+ To align labels center, precede the entry with ^. a caret.

[NOTE: Numbers may be used as Labels when preceded by one of the above positioning symbols.] align

B. Entering Values

1 - Position the cell pointer in the desired cell using the arrow keys.

2 - Type the number then press the **ENTER** key.

[Note: Do not precede the number with an alignment character unless you want the number to be treated as a label.]

C. Entering Formulas

A formula instructs Lotus to calculate the value of a cell from two or more other cells. Formulas may be quite complex and include additions, subtractions, multiplications, and divisions. There a several ways to enter formulas in cells.

The Typing Method

1 - Position the cell pointer in the desired cell using the arrow keys.

2 - Press the + key to begin formula entry.

3 - Type a formula in terms of cell references and then press **ENTER**.

For example, if you want to add the values in cells C3, C5, and C7 and place the result in cell C9, move the cell pointer to cell C9, press the + key, type **C3+C5+C7**, then press the **ENTER** key. Examples of other valid formulas include:

C3*C5 (C3 multiplied by C5)
C5/C7 (C5 divided by C7)
C7-C3 (C7 subtract C3)

+ Lotus also contains a large number of built in functions that make it easier to enter more complex formulas. For example to sum all the cells between C3 and C25 one could enter:

@SUM(C3..C25)

To find the average of the same set of numbers one need only enter:

@AVG(C3..C25)

Many other functions are available and can be found in the LOTUS manual.

The Pointing Method

Two examples of the Pointing Method:

Example 1 : To add the values contained in cells C3 C5 and C7 in cell C9:

1 - Position the cell pointer in cell C9.

2 - Press the + key to begin formula entry.

3 - Move the cell pointer to cell C3, then press the + key. (The cell pointer returns to cell C9.)

4 - Move the cell pointer to cell C5, then press the + key. (Again the cell pointer returns to C9.)

5 - Move the cell pointer to cell C7. This time however **do not** press the + key. The formula is now completed. It is displayed on the second line in the top left corner of the screen and should read:

C3+C5+C7

To enter the formula in the cell, press the **ENTER** key.

Example 2 : To sum the values in all cells between D12 and G12 in cell M12:

1 - Position the cell pointer in cell M12

2 - Press the + key to begin formula entry.

3 - Type **@sum(**

4 - Move the cell pointer to cell D12 with the arrow keys.

5 - Press the . key (Period key)

+ Lotus displays formula as **@sum(D12..D12**

6 - Move cell pointer to cell G12

+ Lotus displays formula as **@sum(D12..G12**

7 - Complete the formula by pressing the) key to close the parentheses.

+ Lotus now displays formula as **@sum(D12..G12)**

8 - Enter the formula into cell M12 by pressing the enter key.

(11) COPYING CELL CONTENTS TO OTHER CELLS

Once you have entered a value, a label, or a formula into a cell it is not difficult to copy the contents of the cell to one or more cells. An example is the easiest way to demonstrate how the copy command works.

Example : In example in (10), above, a formula was developed to add the contents of a range of cells in the same row into a single cell, cell M12. Assume that you wish to add the contents of the corresponding set of cells in rows Q, R, and S and put the respective sums in cells Q12, R12, and S12.

1 - Position the cell pointer in the cell in which the formula you want to copy is located, in this case cell **M12**.

2 - Press the / key to display the Command Menu.

3 - Select **Copy**, then **ENTER**.

+ Lotus displays the following prompt:

Enter the range to copy FROM: M12..M12

+ M12..M12 indicates that the range contains only one cell, M12. In this example you will copy the formula from M12 to three other cells, i.e a three cell range.

4 - Press the **ENTER** key.

+ Lotus displays the following prompt:

Enter the range to copy TO C3

5 - Move the cell pointer to cell Q12.

6 - Press the . key (period key).

7 - Move the cell pointer to cell **S12**.

+ The range of cells, Q12 to S12, will be highlighted.

8 - Press the **ENTER** key to copy the formula from cell M12 to cells Q12, R12, and S12.

The procedure for copying is the same for copying cell values or labels. It is also possible to copy more than one cell at a time.

(12) ADJUSTING COLUMN-WIDTHS

To adjust column-widths within a worksheet:

1 - Position the cell pointer in any cell in the column you want to adjust.

2 - Press the / key.

3 - Select **WORKSHEET.**

4 - Select **COLUMN-WIDTH.**

5 - Select **SET.**

6 - Enter the number of spaces you want the column to be and then **ENTER.**

[NOTE: The "default" setting for Column-widths is 9 spaces.]

(13) COLUMN AND ROW INSERTION

To insert a column or a row, do the following:

1 - Position the cell pointer in the to the location where you want to insert a column or a row.

2 - Press the / key.

3 - Select **WORKSHEET.**

4 - Select **INSERT.**

5 - Select **COLUMN** or **ROW** as desired.

+ Lotus prompts you for an insert range. If you only want to insert the width of one column or row, press the ENTER key. If you want to insert an even greater number of columns or rows, use the ARROW keys to expand the RANGE. After expanding the RANGE, press the ENTER key to enter the RANGE.

(14) PRINTING

To print a worksheet, do the following:

1 - Retrieve the worksheet you want to print.

2 - Prepare the printer.

3 - Press the / key.

4 - Select **PRINT**.

5 - Select **PRINT** again, then press the **ENTER** key.

6 - Select **RANGE**, then press the **ENTER** key.

7 - Enter the Range of cells that you want to print.

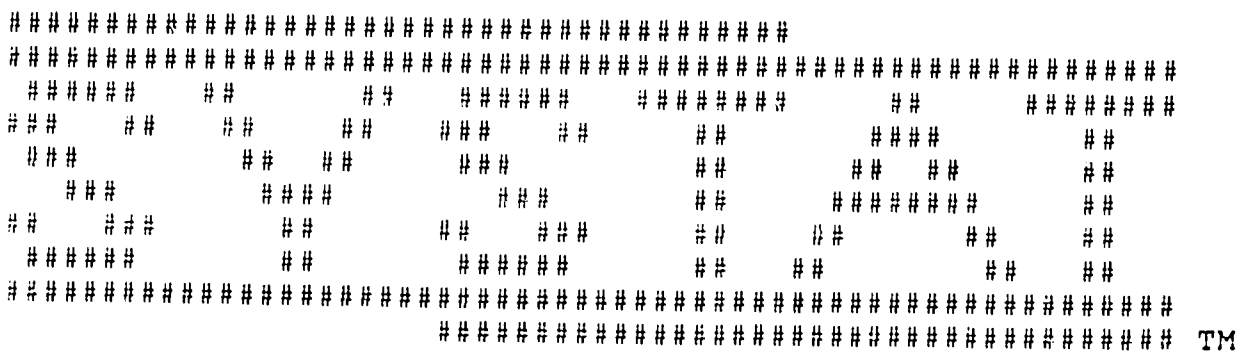
8 - Press the **ENTER** key.

9 - Check the paper alignment in the printer, select **ALIGN**, then press the **ENTER** key.

10 - Select **GO**, then press the **ENTER** key.

[NOTE: Standard paper size permits printing of ranges that are up to 80 characters wide and 66 rows deep. Many printers allow you to modify the print size so that it is possible to print ranges up to 135 characters. With modified print and wide paper it is possible to print ranges as wide as 240 characters. Utility programs also exist that facilitate sideways printing which permits printing of ranges of almost any number of characters.]

GETTING STARTED WITH



GETTING STARTED WITH SYSTAT

PREFACE

The SYSTAT statistical analysis software package has great capabilites. It makes the analytic power previously available only on larger computers available to personal computer users. Although SYSTAT's capabilities are great, it is easy to use. A researcher does not have to comprehend all the package features to begin useful analysis.

This manual is intended to serve as a brief overview. It is intentionally limited in scope. Readers are encouraged to turn to the SYSTAT Manual by Leland Wilkinson for a more complete and thorough explication of the package's capabilites. The first section, Overview of SYSTAT, provides a basic introduction to how the SYSTAT system works. After reading it, you should be able to complete the tasks in Examples 3 and 4 for week three. The second section, entitled Additional Comments on Analysis with SYSTAT, explains how to prepare data and use the analytic capabilities.

GETTING STARTED WITH SYSTAT

SYSTAT can handle large amounts of data and perform extensive analysis using small personal computers. It does so by using a system of modules to manage and analyze data files. With only a brief introduction, a researcher can begin performing useful analysis; the finer points and more complex capabilities can be easily learned along the way. This section includes enough information to get you started using SYSTAT; it will explain;

- o How to start a SYSTAT session,
- o How data files are managed in SYSTAT,
- o How to work with the two main types of modules, and
- o How to use commands to get the analysis you want.

1.1 HOW DO I GET STARTED?

Begin by changing to the directory which contains the SYSTAT program, and type **systat**. When you type '**SYSTAT**' to begin a SYSTAT session, you will first see the screen reproduced below.

```
VERSION 3.0
COPYRIGHT, 1985
SYSTAT, INC.
SERIAL NUMBER IS: 7374

THIS PROGRAM BELONGS TO William Cummings, Harvard, Cambridge, MA
PLEASE CALL (312)-864-5670 IF YOU FIND AN ILLEGAL COPY


Press RETURN to go to the menu.
```

Hit **<Enter>** to procede to the main menu. There, you will see the options listed in the screen reproduced on the next page.

```
SYSTAT 3.0  William Cummings, Harvard, Cambridge, MA                SER  7374
   1 DATA      2 GRAPH     3 STATS     4 TABLES    5 NPAR      6 CORR
   7 MGLH      8 FACTOR    9 MDS      10 CLUSTER  11 SERIES   12 NONLIN


Enter number, module name, HELP, or other command.  QUIT returns you to DOS.
```

Your first choice in a SYSTAT session is which module to use. These are explained in the following subsection.

1.2 HOW DOES SYSTAT HANDLE DATA IN FILES?

There are two types of modules in the SYSTAT system: data management modules, and analytic procedures modules. The data management modules (option 1 on the screen shown above) take in data, code it in such a way that the other modules can read it, and store it. The analytic modules (options 2 through 12 in the screen above) read the data from the files, perform whatever statistical analysis you desire, and show you the results, either on the computer's screen, in a printout, or by recording them in an output file which you can examine later. A brief description of the three components--data files, data management modules, and analysis procedure modules--follows. For more detailed descriptions consult the SYSTAT manual.

1.21 SYSTAT Data files

Data management in SYSTAT differs from data management in spreadsheet programs in a number of ways. Structurally, instead of pulling the data onto a worksheet and letting you work with it, SYSTAT holds the data out of sight and takes commands from you through the different modules. In terms of data storage, SYSTAT files are similar to spreadsheet files in that they contain columns of numbers; each row constitutes a record, in our case, the data for a particular

school. They differ, however, in that SYSTAT is more rigid about the data it can contain. It is like a spreadsheet with locking sides; if you open a new column, SYSTAT needs something to put in each row of the column; only complete rows or columns of data can be entered. One cannot, for example, move to the side of the data and construct tables; SYSTAT files contain only data, no formulas or tables.

All SYSTAT file names have two parts. You provide the first portion, and SYSTAT provides the last part. You can choose any name 1 to 8 characters in length. Names can contain numbers or letters, but no special symbols; they must begin with a letter. It is best to use names that describe what is in the file and that will help you remember the files' contents. For example, we have named the three SYSTAT data files used in the workshop as follows:

CENDATA contains data from the school census,
PRSDATA contains data from the principal's survey,
PRQDATA contains data from the principal's questionnaire.

The second part of the file, the three letter extension, is added by SYSTAT. It identifies the file type. For example, all of our three example files are data files; SYSTAT assigns them the extension .SYS. Thus, the full names for these files are CENDATA.SYS, PRSDATA.SYS, and PRQDATA.SYS. The other file extensions assigned by SYSTAT are .DAT, which signifies files which can be shared between SYSTAT and other software packages, and .CMD, which identifies files containing records of commands. We will not deal directly with these types of files. Files appear in your disk's diretory by the full names; within SYSTAT, you need only refer to them by the first portion as in the command;

>use cendata <Enter>
>get cendata <Enter> for a .DAT file that is being sent from LOTUS, Rbase, or another data handling program.

1.3 HOW DO I WORK IN SYSTAT MODULES?

The main menu in SYSTAT gives you a choice of modules to work in. The first module listed, the Data module, enables you to enter and prepare data. The remaining modules allow you to perform various kinds of analysis. Both the data and analytic modules are used throughout the research process as data are entered, analyzed variable by variable (univariate analysis) in pairs or groups, as indicators and ratios are created and further multivariate analysis is conducted.

1.31 Data Management Modules

DATA module and the EDIT facility within that model allow you to enter and change data. These are the basic processes in preparing the data for analysis. Initially, if the data are not in a SYSTAT file, they must be entered. Once the data are in usable files, additional changes, additions, calculations may be desired prior to analysis. For example, a researcher might need to merge the variables from a number of files into one file, or transform or create new variables, for example, by adding existing variables together.

1.32 Analytic Procedures Modules

The second step in using SYSTAT to perform analysis is to use the procedure modules' commands to create analysis of the data. The procedures modules are capable of a great variety of sorts of analysis; most procedures modules can also analyze subgroups of the data. This is useful in comparing groups of schools, etc. The modules enable you to perform analysis of all sorts, from examination of frequency tables to correlations and regression. Eleven different procedures modules are available.

Records of the analysis, the output of the procedures, comes to the screen unless SYSTAT is instructed otherwise. You can choose to send the output to a printer or to a file for later printing. These options are explained below in the section on data analysis.

1.4 HOW DO I GIVE SYSTAT COMMANDS?

You will recall that in LOTUS as in many other software packages, commands are entered by choosing an option from a menu or a command line. In SYSTAT modules, commands are given by typing the names of commands at the prompt, which looks like this >. While the choices of commands are not displayed across the screenn as in LOTUS, they are limited in number. To see which commands are available to you at any point in your SYSTAT session, type help, followed by a return (which might be labelled 'Enter' on your keyboard), at the prompt, like this:

>help <Enter>

SYSTAT will then display the commands or options available to you. Usually, additional specific information regarding the commands is available by typing 'help' followed by the command name, for example, typing

>help save <Enter>

will bring up a help screen describing the use of the 'Save' command.

Commands given to SYSTAT are always followed by a <Enter> or <Enter>. SYSTAT will not procede without the <Enter>. There are two main types of commands in SYSTAT, **!HOT!** commands, and **COLD** commands. Most of SYSTAT's commands are **COLD** commands; this means you are required to type **RUN**, followed by a <Enter>, before anything will happen -- just as you would in writing small computer programs. In using SYSTAT to perform analysis, your commands will often form a small program. The series of commands will be started either by typing **RUN** or by typing a **!HOT!** command. In the analytic procedures models, you will most often need to specify conditions for the analysis, and conclude the list of statements with a !HOT! command that makes the analysis take place. A few of the important commands you'll want to use right away are described below.

1.41 Commands in the Data module

These are some of the common commands you will use in the Data module to combine files, create and transform variables, and save the new data in forms that can be read by the analytic procedures modules. The commands are: USE, GET, LET, RUN, SAVE, PUT, IF...THEN and ELSE, and SORT.

USE This command, followed by a file name, tells SYSTAT which file you want to examine or alter. You can select one whole file, combine two whole files, or choose some of the variables form one or two files to pull together into one file. If you are pulling information from two files, the data must be ordered the same. For example, two files containing information from the same three schools can be read together as long as the schools are ordered the same in the two files.

To use one file, enter the following using your file name:

>use filename

To use specific variables from one file, enter your file name with the variable names in parentheses, separated by spaces or commas. For example,

>use filename(var1 var2 var3)

To combine variables from two files, enter the equivalent of the following.

>use file1(var1 var2 var3) file2(var4 var5 var6)

Note that the filenames and variable names must be different. If you have variables of the same name in two different files, DO NOT READ THEM INTO THE SAME FILE. This will cause great confusion.

ALL SYSTAT FILES SHOULD CONTAIN AN IDENTIFYING VARIABLE SO THAT THE ORDER OF CASES CAN BE INSPECTED. FOR EXAMPLE, EACH OF OUR DATA FILES CONTAINS THE VARIABLE 'SCNO', THE SCHOOL'S IDENTIFYING NUMBER, SO THAT WE CAN CHECK THE ARRANGEMENT OF THE DATA AT ANY TIME.

==========

GET This command is similar to the USE command; it is used to 'get' data from non-SYSTAT ASCII files. When you tell SYSTAT to 'GET FILENAME' it looks for a file by the specified name with the extension **.DAT** , and reads in the data. Additional comments on reading in data in this form are included in section 2 below.

==========

LET This command is used to assign values to a new or old variable. For example, if you have a file with information from schools which contains a variable STUDENT which records the number of students, and a variable TEACHER which records the number of teacters, and you wish to compare the student-teacher ratios of the school, you could issue the following command to SYSTAT:

>LET RATIO=STUDENT/TEACHER

SYSTAT will first check to see that it understands what you want it to do, then wait for additional commands until it receives the **run** command. It will then create a new variable for each case (in this example, for each school in your data file), and assign it the value of the number of students divided by the number of teachers. You can perform

all possibe mathematical operations using variable names or numbers in creating or transforming variables.
Note: In performing transformations, SYSTAT will attempt any possible mathematical operation; it will not, however, do the impossible. It will not, for example, divide by 0. If a researcher attempts to calculate the ratio of students to science laboratories in a group of schools using the command **LET LABRATIO=STUDENT/LABS** , and several schools have 0 science laboratories, SYSTAT would abort the program when it reached the first school where LABS=0. **Check your variables carefully to be sure you're not asking SYSTAT to do the impossible.**

==========

RUN This command tells SYSTAT to execute all of the 'cold' commands you have given it. If you are making changes in a file but fail to specify a new file to save the data to, SYSTAT will give you a warning message.

==========

SAVE This command, followed by a filename, tells SYSTAT to save the data you are working with, including any new transformation, into a file. SYSTAT's first response to this command is to check the directory to see if the name has been used already; if it finds a file with the name you've chosen, it will ask if you want to write over the file. If you respond 'yes' (by typing **y** at the prompt), SYSTAT will write over the previously saved file, and that data will be lost. If you respond 'no' (by typing **n** at the prompt), SYSTAT will expect a new **save** command with a new file name assignment.

==========

PUT This command is similar to the SAVE command except that it directs SYSTAT to store the data in a form that can be read by other software packages such as LOTUS or RBase. The command **PUT FILENAME** tells SYSTAT to store the data in an ASCII file with the extension **.DAT.**

==========

IF .. THEN , and ELSE These 'logical operators' allow you to perform more complex transformations of data. Let us say a researcher wanted to divide a sample of schools into 3 groups based on the size of their faculty. Using the variable 'faculty', the researcher could create variable for 'size' as follows:

```
>10 if faculty=0 then let size=0
>20 if faculty>0 and faculty<10 then let size=1
>30 if faculty>10 and faculty<20 then let size=2
>40 if faculty>20 then let size=3
>save data5
>run
```

Notice from this example that SYSTAT allows you to write small basic programs in the Data module using numbered lines; the logical operators 'AND' and 'OR' are also available to you in transforming data in the Data module. A longer example is included in Example 4 of your materials, and further information is available in the SYSTAT manual. Note too that the same limitations mentioned under the **LET** command apply here -- SYSTAT can only do the possible, and will abort attempts do the mathematically impossible.

The **ELSE** command comes on a line of its own following an IF..THEN statement. It is followed by an expression which is executed only if the preceding IF..THEN statement is not executed. For example, to create a dichotomous variable which had the value 0 for small faculties, 1 for 'big' faculties, we could use the following program:

```
>if faculty>20 then let big=1
>else let big=0
>save data6
>run
```

This program would create a variable called BIG which would contain the value 1 if the faculty was larger than 20, but would contain elsewise contain the value 0 for faculties of 20 or less.

==========

SORT Finally, in order to do certain kinds of analysis in the other modules, Data must be sorted according to the groups the researcher wants to compare. To carry through the last example, if the researcher wants to later compare the schools in groups based on faculty size, the following program could be entered.

```
>use data5
>sort size
>save data6
>run
```

SYSTAT would then sort the data according to the variable 'size', and save all of the data in a new sorted file.

==========

Additional information about commands in the Data module are available in the SYSTAT manual or by typing **help** in the data module. Additional information about specific commands can be obtained by typing **help** followed by the name of the command.

1.5 COMMANDS IN THE ANALYTIC PROCEDURES MODULES

Each of the modules contains specific commands for doing particular analyses. This section contains information about commands and practices that are common to many or all of the modules. For more detailed answers to specific questions, use the on-line help facility or consult the SYSTAT manual. The commands are: BY, OUTPUT, PRINT, and SELECT.

BY This command is entered, followed by a variable name (or 2 variable names), to produce analysis by groups. For example, in generating summary statistics for our schools in the example above, we could give the following commands in the Statistics module:

```
>use data6
>output @
>by size
>statistics var1 var2 var3
```

This would produce a print out of the separate summary statistics of the variables listed for the groups of schools having the same value for the variable 'size'. Note that you needed to use a data file taht was sorted by the variable specified in the 'BY' command.

==========

OUTPUT This command tells SYSTAT where you want it to send the results of the analysis. Unless you specify otherwise, SYSTAT will send output to the screen. In most cases, howver, you will want a hard copy of your analysis. Entering the command **output @** sends output to the printer. To change back to the screen, enter the command **output ***.

==========

PRINT This command allows you to request greater detail in your output. In most analytic procedures modules, you can specify **print=long** to obtain additional pieces of output from the analysis. To change back to the default setting, enter the command **print=short.** Check the details for the particular analysis in the SYSTAT manual to decide if you need the additional information; in many cases you will not.

==========

SELECT This command allows you to specify only specific cases which meet some criterion for analysis. For example, if you wanted to produce analysis <u>only</u> for the schools with larger faculties, you could enter the command sequence:

```
>use data5
>select size=3
>statistics var1 var2 var3
```

Unlike the 'by' command, the 'select' command does not require that the data be sorted. It looks through the whole file to pull out those cases which meet your specification.

==========

There are many commands within each particular modules, and often options within those commands. For general information about the individual modules, type **help** followed by the module name in the main menu. For information about the commands available in a particular module, enter that module and type help. For more specific information regarding those commands and the options within them, type **help** followed by the command name within the module, or consult the SYSTAT manual.

ADDITIONAL CONSIDERATIONS IN ANALYSIS WITH SYSTAT

2.1 DATA PREPARATION

The first step in using SYSTAT in analysis is to prepare the data. Data preparation is accomplished in the Data module of the the package; enter Data by typing 1 <return> at the prompt. Two stages of data preparation are dealt with in this section; the first is to enter the data into a file that can be read by the program. Data can be entered directly into a SYSTAT file, or can be read from a file created by another data handling software package such as LOTUS 1-2-3 or RBase. Both of these methods are discussed below.

Once the data are entered into a readable file, additional preparation may be necessary. In the data module, SYSTAT allows variables from different files to be merged into a single file, multiple variables that describe the same phenomenon can be combined into indicators, or other transformations that you may desire such as the creation of attendance rates from records of days absent can also be performed.

2.11 Entering data into SYSTAT files

Data can be entered directly into a SYSTAT file using the edit facility in the data module. In the Data module, type **Edit** at the prompt. The screen reproduced below will appear with a highlighted border. The cell to the right of the word case will be highlighted. The names of your variables are entered in the top row; data entries begin with the row labeled 'Case 1'.

insert editor screen

Variable names can be 1 to 8 characters long, and may contain letters and numbers. They must begin with a letter. In making initial entries, names of variables must be enclosed quotation marks or apostrophes. When you enter the name of a variable, SYSTAT will assume the data beneath that variable will consist only of numbers **unless the final character is the '$'**. For example, to create a file containing data from a number of schools, an appropriate name for the variable 'school name' would be **'SCHOLNM$'**.

Non-numeric data must be enclosed in quotes when you type it in, just as all variable names. SYSTAT will respond with an error message if you try to enter numbers where it expects a character string, or if you try to enter a word or other character string where it expects a number. If you receive such an error message, check the variable name to be sure which form of input SYSTAT was expecting.

Variable names and data can be entered in the 'window' that appears when you call up the editor from the Data module. In order to save your file or to enter additional commands available in the Edit mode, you must move the cursor to the command line below the window. This is done by pressing the **Esc** key (or its alternate, Ctrl-Q). This moves the cursor to the command line; pressing it again will return the cursor to the edit window. From the command line, you can type the command **SAVE**, followed by the desired filename, in order to save your data. For example, if you had just entered data from a sample of schools in a district, you would leave the editing window by typing the 'Escape' key (<Esc>), and would type the command to save as follows.

>SAVE SCHLDATA <Enter>

SYSTAT would then check your disks directory to see if that name had been used for another data file. If the name is already used, SYSTAT will display a warning message like the one below asking if you want to write over the existing file.

INSERT SCREEN DUMP HERE

453

If you answer 'yes' by typing a **y** (followed by <Enter>), SYSTAT write the new file on top of the old one; the old data will be lost. If you answer 'no' by typing **n** (followed by <Enter>), SYSTAT will return the prompt and wait for another save statement. If the file name is not already used, SYSTAT will save the file and return the command prompt. You can exit from the editor by typing **quit**. If you fail to save the file before attempting to quit, SYSTAT will warn you and ask for a filename to save the data.

Data can also be read into a new SYSTAT file without use of the Edit facility. First, a name is created for the new file using the **save** command, then the **input** command is entered, followed by the names of the variables to be entered, for example;

```
>save schldata
>input school$ district enroll <Enter>
```

Following this, data can be entered, separated by a space or a comma. **Each line of data entered must contain the right information in the right order.** Thus, the following data could be entered:

```
>Jefferson 732 456
>Lincoln 730 657
>Kennedy 730 413
>MLKing 732 536
```

Data entry is completed by typing a word if SYSTAT expects a number, or by typing the ~ if SYSTAT expects a name. In our example, SYSTAT expects the next data line to contain a school's name, so entry would be terminated by entering a ~;

Data can also be read into SYSTAT from data files created on another software package such as LOTUS, RBase, or others, using the **GET** command. These files must have the file extension **.DAT** in order to be read. These files may only contain the data, and the variable names must be entered using the input command. For example, the following

program would bring in the same data as the previous one if the data had been entered into a LOTUS file which we renamed SCHOOLS.DAT.

```
>save schldata
>input school$ district enroll
>get schools
```

2.12 Data Cleaning

Data must be handled carefully throughout the research process. Two particular areas of concern are the newly entered data, and data that are 'created' by combining or transforming original data. Both of these areas are discussed below.

Before analyzing the relationships between and among variables, the researcher must examine the individual variables carefully. Data can become 'contaminated' in a number of ways at all stages of the process, including the following:

A. In data gathering, incomplete or unclear recording can lead to loss of data. It is important to take care in instructing data-gathering personnel to code all information clearly, e.g. coding 'no' answers as 'no', nnot merely leaving a blank. **Missing data cannot be analyzed!**

B. In data entry, data can be misread so the wrong value is entered, misplaced so that values entered in the wrong column, or mistyped. In all cases, information is lost. **It is essential to keep all original data records for** later reference!

C. During analysis, disks can go bad, editing can accidentally destroy data, or other problems can occur. Computers are a great aid to analysis, but they are not infallible. **Always keep back up copies of data!**

Prior to extensive analysis, careful cleaning of the data is needed. This includes:

A. Examining the frequency distribution of the variables. If a variable is to be coded 0 or 1, and several cases show values of 5, the researcher knows there is a problem. Referring to the original data records is the best

recourse at this stage. In some cases, two variables may be transposed, for example, if the number 1988 occurs in the variable 'MONTH' while the number 7 occurs in the column for 'YEAR'.

B. Checking the number of missing values recorded for each variable. Depending on the nature of the information coded, this has different ramifications for the analysis. If there is a great deal of missing data, the researcher is limited in using the variable in combination with others to create indicators.

2.2 DATA TRANSFORMATION: INDICATORS, RATES, CATEGORIES

Once the original data are entered and cleaned, another set of concerns confront the researcher as the creation of indicators and transformation of data is begun. Three common operations with data, creating indicators, calculating rates, and recoding variables to different levels of measurement, all have particular concerns.

2.21 Indicator Creation

Often, when several pieces of data are combined to form an indicator of a broader concept. For example, socioeconomic status is usually measured by education level and by income level. Thus, if researchers have collected data on individual levels of income and education, they may want to combine the variables to create an indicator of socioeconomic status. Several steps should precede the creation of the indicator.

A. Univariate analysis of each of the pieces to be combined should be carried out. Ideally, both the income and education variables would be somewhat 'normally' distributed; if everyone's income or education level was the same, then adding the information together would not reveal anything. In SYSTAT this is accomplished using one of two approaches. For catagorical data (data coded so it's possible values are a limited number of integers), the TABLES module is used to generate a frequency distribution. The command is

>TABULATE VARIABLE/LIST

For data that are continuous (variables such as age or weight, where the scale is clearly an interval one) the more appropriate approach is to generate desriptive statistics of the distribution in the STATISTICS module. The command is:

>STATISTICS VARIABLE

Note that for a continuous variable like age where the subjects age is rounded off to the year, all data values

will be discrete, so a frequency distribution might also be revealing.

B. Correlation between the pieces should be calculated. In SYSTAT this is done in the CORR module with the PEARSON command. If the correlation between two variables is very low, it makes little to combine them since they are probably not indicating the same thing. If the correlation between them is negative, they will cancel each otehr out. Pay careful attention to the direction of the scaling for each variable. In our workshop data, the variables for education level and income were coded in opposite directions; both ranged from 1 to 5, but for one of them, 1 represented the highest level, for the other, the lowest.

As noted in the earlier section describing the **let** command, particular care must be taken in adding or multiplying variables. SYSTAT will not divide by zero; all calculations which include a missing value will be coded missing (e.g. adding two variables where one has a value and the other is missing will produce a 'missing' sum). These factors must be considered in creating indicators, because large numbers of missing values in constituent variables could result in a wonderful indicator that is coded as 'missing' for all but a few cases.

C. Attention must be given to missing values and zeros. Bear in mind that SYSTAT cannot perform mathematical calculations that include missing values. If a value used in an equation is missing, the result of the equation will be coded as missing. If one, both, or several of the variables you will combine to form an indicator are missing, the resulting indicator will be mising for each case where any one portion is missing. Also remember that SYSTAT cannot perform impossible mathematical calculations. If a zero appears in the denominator, SYSTAT will abort the program and save a partial file.

2.22 Calculation of Rates and Ratios

It is very easy to calculate rates in SYSTAT. This is the situation where the presence of zeros in the variable which forms the denominator of the equation poses a great problem. To complete the calculation of rates for all cases which have non-zero values, two strategies are possible;

sample programs are shown on the next page.

The first approach is to create a modified version of the variable which is the denominator, substituting the . which codes missing cases for the zeros. For example, if we were calculating the ratio of students to science labs in a group of schools, some of which had no science labs, we could create a modified version of the LABS variable. Where the variable ENROLL contained the number of students and the variable LABS equals the number of science labs in a school, such a program would look like this:

```
>USE SCHLDATA

(HERE SYSTAT WOULD LIST THE VARIABLES AVAILABLE)

>LET NLABS=LABS
>IF LABS=0 THEN LET NLABS=.
>LET LABRAT=ENROLL/NLABS
>SAVE LABDATA
>RUN
```

The second approach uses the IF..THEN capabilites to avoid the cases where the denominator equals zero. For example:

```
>USE SCHLDATA

(HERE SYSTAT WOULD LIST THE VARIABLES AVAILABLE)

>IF LABS=0 THEN LET LABRAT=.
>ELSE LET LABRAT=ENROLL/LABS
>SAVE LABDATA
>RUN
```

As in the case of creating indicators, all calculations which include missing values result in missing values.

2.22 Categorizing Data

Using the IF..THEN capabilites of SYSTAT, data which are coded as continuous can be transformed into categorical form. See the example under the IF..THEN, and ELSE command in the earlier section. If the underlying distribution of a variable is continuous but it is coded in discrete

categories on a different scale, an approximation of the underlying continuous distribution can be produced.

2.3 ADDITIONAL COMMENTS ON THE ANALYTIC MODULES

Using the **HELP** facility, you will find the analytic modules to be fairly self-explanatory. The list below reviews the major functions for which you might use the SYSTAT modules to perform analysis for a policy study.

GRAPH This module allows you to create plots of single or multiple variables. These include scatterplots, two-way plots, bar graphs of frequencies, and histograms of distributions. These functions are very useful in

describing visually the distribution if single variables and the relationships between variables.

STATISTICS This module produces descriptive statistics of variables. It is important in the initial stages of cleaning variables and examining univariate characteristics of the data. It is also useful throughout analysis in comparing differences between groups, etc.

TABLES This module produces tables and cross-tabulations of variables. In the initial stages of analysis it is used to produce frequency distribution tables. It is useful in producing crosstabulations.

CORR (Correlation) This module produces correlation statistics. In addition to other uses, it is used in examining the relationship between/among variables which might be combined into indicators.

MGLH (Multivariate General Linear Hypothesis) This module produces simple regression and other multivariate models.

459