DOCUMENT RESUME

ED 115 855 CE 005 693

AUTHOR Breslow, Martin P.
TITLE A Very Tentative Computer System Model. Occasional Paper No. 3.
INSTITUTION Management and Information System for Occupational Education, Winchester, Mass.
SPONS AGENCY Massachusetts State Dept. of Education, Boston. Div. of Occupational Education.
PUB DATE 14 Feb 72
NOTE 35p.; For related documents, see ED 062 553; ED 068 646-647; ED 072 225; ED 072 228; ED 072 303-304; CE 005 687-727; Figure 11 will not reproduce in microfiche due to small size of type

EDRS PRICE MF-$0.76 HC-$1.95 Plus Postage
DESCRIPTORS *Computer Programs; Data Analysis; Data Collection; Data Processing; Decision Making; Demonstration Projects; Electronic Data Processing; Information Processing; *Management Information Systems; *Program Development; *Simulation; State Programs; Systems Development; *Vocational Education
IDENTIFIERS *Management Information System Occupational Educa; MISOE

ABSTRACT

The developmental paper, one of a series written as the Management Information System for Occupational Education (MISOE) was conceptualized, is a first attempt to picture the computer system necessary to carry out the project's goals. It describes the basic structure and the anticipated strategies of development of the computer system to be used. Three major subsystems are discussed: the data entry subsystem (processing raw census and sample data organized into input, process, product, and impact data types), the analysis subsystem (performing second level and contingency analyses), and the optimization and forecasting subsystem (providing decision-making tools through mathematical and simulative analyses). The use of computer simulation as a decision-making tool is discussed in some depth. System development in two phases is outlined. All figures supplementing the text are placed in the appendix. (Author/MS)

ED115855

OCCASIONAL PAPER #3

A VERY TENTATIVE COMPUTER SYSTEM MODEL

Martin P. Breslow

February 14, 1972

Management and Information System for Occupational Education
1017 Main Street
Winchester, Massachusetts 01890
Telephone: 617-729-9260

TABLE OF CONTENTS

PREFACE

This Occasional Paper is a first attempt to picture the computer system necessary to carry out the goals of MISOE. By the end of June of this year, we plan to offer a much less tentative description than is presented here. This paper draws on Occasional Papers #1, #2 and #4 for supplement and the reader may want to have them handy. There is also a brief reference to Planning Chart No. 1.

The nature of the data base, the kinds of analyses we will use, the decision making tools we will provide and the stages of development we anticipate, are all discussed. The use of computer simulation as a decision making tool is discussed in some depth. In it is the opinion of the MISOE staff that the simulation technique captures the essence of the kind of system we plan to offer.

Finally, I would like to call attention to the placement of all Figures in the Appendix. I felt that including them in the text would require too much turning of the pages and this would disturb the continuity of thought I have tried to maintain.

TABLE OF FIGURES

FIGURE

This paper describes the basic structure and the anticipated stages of development of the computer system we will use. The approach taken is functional. That is, what kinds of things the system will be able to do are discussed in some detail. Only a limited amount of space will be devoted to the "hows" of the system. First we will discuss the three major sub-systems; namely, the Data Entry Sub-system, the Analysis Sub-system, and the Optimization and Forecasting Sub-system.

I. THE DATA-ENTRY SUB-SYSTEM

As implied in its name, this sub-system's function is to process raw census and sample data as it enters the system on punched cards and, perhaps, occasionally on magnetic tape. The data entry sub-system is organized by data type, i.e., by input, process, product and impact data types. Figure I depicts the processing performed on (school system) input data. The data having been classified as input data is further classified by major topic (student type or expenditure) and by source (census or sample) and, of course, the data will have already been coded by stratification level as it comes in.

The results of this surgery are maintained while totals, averages and whatever other low-level descriptives might be required are accumulated. The nature of these other descriptives will be determined by what subsequent analyses will be performed in the other sub-systems. When this initial processing has been completed, the results will be printed out in the form of a "pre-determined" set of reports. By "pre-determined", it is meant that any changes (e.g., additions and deletions of reports) can be made only by someone with intimate knowledge and direct access to the system.

All data is stored according to two classificatory schemes. One is by the kind and the location of the <u>decision maker</u> whose acts are reflected in the data. The other is by who is directly involved in the situation described by the data. Occasional Paper #4 by Elizabeth Weinberger details the two schemes for process data. An example might be a local school administrator's decisions on class size - Scheme 1-- and this can be alternately classified as an assignment of students (the individuals directly involved) to a physical factor --Scheme 2.

Ms. Weinberger suggests a numbering system in which the first four numbers of the relative storage address of a piece of data refer to the "role incumbent" who made the decision, or the role incumbent who was involved in the data situation. The remaining (3 or more) numbers refer to the data's classification in its related space (input, process, product or impact). These addresses are those of a magnetic disk. This peripheral storage device is addressable in a manner similar to the way the computer's memory is addressable. One can go directly to specific locations to pick up data. The unit of storage is, however, larger in the disk than it is in the computer. On the disk, the unit of storage is always several characters in length and it is called a "record". For convenience sake, one often works with addresses relative to some starting point rather than the absolute disk address.

In the example, the suggested Scheme 1 relative address is 4002131 and the Scheme 2 relative address is 4003360. The numbers 4003 refer to a local within a school administrator (location-school, kind-administrator) and the numbers 4002 refer to students in a school (location-school, kind-student). The last three numbers of the first, 131, refer to a decision-making behavior. The last three numbers of the second refer to an organization of human factors with a physical factor. The addresses XXXX0000 to XXXX0370 refer to process data. The addresses XXXX0371 to XXXX0740 might refer to input data, and so forth.

Now, while the data is stored according to type (I,P,P,I), most analysis routines require that we retrieve data, connected by student across types. In addition, we may want to go from one scheme to the other. For example, looking at a given decision maker, we may find out that this role incumbent determines room size in schools A, B, C but not D, E, F. We might want to enter Scheme 2 at the appropriate address for room size where there is along with individual room sizes in individual programs, the code number of the role incumbent who chose the room size in each case. From this latter data, we could determine who were the decision makers in schools D, E and F. If we are going to be able to make both of these two kinds of connections, at some later time (when data is retrieved from storage for analysis), we must establish linkages prior to the initial storage of the required data items.

In Figure 1, the formation of connecting links follows the reporting function. Figure 2 provides a simplified example of what the connecting links across data types might look like. Each strip is a Scheme 2 record on a magnetic disk. The records illustrated in Figure 2 are organized as follows: The first part contains data of a given type, say, input, on a given student; the other parts contain the addresses at which that student's process, product and impact data are located. Using the first address code (100), we can retrieve the record containing that student's process data. This process data record, of course, contains process data in its first part. It also contains addresses but its first address does not point to process data; it points to input data. Hence, a program can use the input record to locate process data and vice versa.

Once these connecting links have been formed, then the data can be stored, freeing computer memory and terminating this data entry sequence. Figures 3, 4 and 5 show the data entry sequences for process, product and impact data, respectively. Note that these sequences differ only in the kinds of connecting links established. Of course, each sequence will be specialized for its own data formats but this is merely detail to be worked out when we are further along.

Naturally, there will be more than one datum per type for each student and each datum will have several connectors of each type (e.g., several product links to a process datum and vice versa). As new data comes in the number of links increase. The links between schemes will also be in the form of addresses stored with data items. Links of different types will be arranged in a fixed order so that they may be identified by programs performing analyses. Through the use of connecting links, we obtain alternate data organizations without duplicating storage.

Not shown in any of these figures is a Data Accounting function which operates on all sub-systems. In the Data Entry sub-system, its purpose is to maintain a map of the data base. This map keeps track of the contents of the reserved data areas, i.e., empty, hypothetical data or real data. There is also a map of programs and routines. Together these maps constitute an index of the system. Later it will be suggested how an index may be used to add flexibility to the systems' operations.

II. THE ANALYSIS SUBSYSTEM

This sub-system performs second level and contingency analyses. Initially, the set of analyses will be pre-determined. At a later phase of development, this constraint will be relaxed and the set will be variable.

Figure 6 illustrates an example of second-stage and contingency process-product analysis. As part of process-product analysis, a multiple-correlation is performed between user determined product and process measures. Following this the contingency of a 50 per cent of variance accounted by process variables is evaluated. If the condition is positive, a finer analysis is performed. Each process variable is examined for the amount of variance it contributes and depending on whether a variable does or does not meet variance and significance criteria, the variable is classified as important or unimportant.

The results of all analyses are printed out in the form of reports and results and useful sub-results are stored in a product file. Each analysis classification has its own results and sub-results files located near the related data file. If we select one ordering of analysis types A_1 - A_6 (see Planning Chart No. 1) as prime, the programs performing analyses can be written with the option of using the previous programs' products so stored to save computation time. Of course, system flexibility requires that most programs also be able to perform these calculations for themselves. Programs looking at longitudinal series of results of the previous analysis must make use of the products of those analyses. Hence, the storage of previous results is itself not an option but a necessity.

Figure 7 illustrates an example in which a product-cost analysis uses the important/unimportant classification of the process-product analysis; it provides information on their comparative influence on cost. Not that this comparison is made or is not made contingent on the determination that the selected process variables taken together significantly influence cost of the product. By combining the two kinds of analyses, we obtain the two-way classification shown in Figure 15. This sort of information should prove very useful in making decisions on the desirability of alternate processes from a cost benefit point of view.

Our ability to make use of sub-results and final results depends, of course, on our ability to detect commonalities among analyses. This suggests the value of maintaining as much uniformity in notation as possible.

Another reason for wanting to retain sub-results is that they may be interesting in their own right. At some stage, we will want catalog them and make them available on user request.

III. THE OPTIMIZATION AND FORECASTING SUBSYSTEM

The Subsystem

The function of this subsystem is to provide techniques for answering questions such as "What is the best mix of...?" and, "What will happen if we do...?" Usually, questions about the best are tempered by such constraints as "The best we can afford,"..."The best results with the minimum change". A number of mathematical techniques are available from curve-fitting to linear programming, queue theory, etc. The advantage of mathematical techniques is that they provide an explicit procedure for finding the optimum solution. Furthermore, many canned programs are available which are highly efficient in terms of computer time and computer storage space.

All these advantages accrue to the user if his system fits the constraints of the mathematical model. If not, the user must decide whether or not he wants to invest in an attempt to develop a mathematical model which can handle his system. The greater the complexity of the system, the fewer the constraints on the stability of relationships, and the greater the number of discontinuities in system process, the less chance one has in being able to develop a successful mathematical model.

An alternative approach which can handle complex, changing, discrete systems is simulation. The optimizing and forecasting subsystem should contain the ability to perform both mathematical and simulative analyses.

Furthermore, it should be provided with data, constants, rates and functions it needs by the other two sub-systems. It would also be desirable to provide this sub-system with the capacity to evaluate its own predictions against past data and, periodically, against new data, as it comes in. This evaluation would be in terms of pure accuracy, and in terms of cost of inaccuracy to the information user.

Simulation: Emshoff and Sisson classify simulation as a "procedural model". And a procedural model is one that:

> "expresses the dynamic relationships that are hypothesized to exist in the real situations by means of a series of elementary operations on the appropriate variables. These operations are stated usually on a computer-like flow chart...[or in decision tables]. The prediction of outcomes is made by actually executing the procedural steps with the appropriate initial data and parameters."

One can also mix mathematical modeling with simulation modeling. For instance, one can generate some of the simulation parameters by randomizing and statistical techniques.

An example of a simulation flow chart appears in Figure 9. It exhibits many processes that are common to all simulations (see Figure 13). Here, levels are numerical amounts or quantities, and rates are rates of change of levels. The model used here is stochastic. Estimates of statistical parameters, including variance estimates that are obtained from the analysis sub-system, might be used here. The simulation will have to be run several times to obtain the raw data needed to compute estimates of outcome parameters. One way of reducing the number of runs needed would be to use expected values in a deterministic fashion ignoring measures of dispersion. Another alternative is to limit stochastic modeling to these variables having the most sensitive effects on outcome values.

Another feature of our example is that it would be designed to handle change. One begins with the system in a steady state. Then there is a transient period when change is introduced and, gradually, the system returns to a new equilibrium (hopefully) or steady state reflecting the long run effect of the change. In Figure 9, the time line is compressed into three stages. Realistically, one would most likely want to spread the time line out in order to make finer observations on the effects of introducing this change. In particular, one would want to know if there are any short term extreme effects and how long do they last before the system stabilizes.

Calculations are those required by the model. The model contains procedural steps and contingency tables to describe discrete processes, e.g., an administrative decision to do x, y or z based on contingencies a, b or c. It contains a set of equations for describing continuous relations. And, there are a set of procedures which tell which values in which equations have to be re-evaluated given a change in any one equation. Figure 10 lists a partial set of equations for a simulation of the interaction of the Occupational System with the World System at input and impact junctions.

The simulation model is diagrammed in Figure 11. The symbolism is derived from that used by Jay W. Forrester in his World Dynamics (1971). Figure 14 will be helpful in describing the elements of this symbolism. Rectangles represent levels or amounts, e.g., number of people employed during the current simulation interval. Octagons represent rates or rate of change functions, e.g., employment rate or birth rate. And, circles will represent "multipliers" of rates. In our example, a sub-system of the World System (see Figure 11) is analagous to an impact as described in Occasional Paper #2 by Dr. Conroy. Each sub-system consists of a current level and a rate of change (a rectangle and an octagon).

The two sub-systems in Figure 14 interact with each other (in a loop) by exerting a multiplicative (interger or fractional) effect on each other's rates.

In Figure 11, we depict some of the possible sub-systems that might appear in a large simulation. We might note in passing that simulation as global as that of Figure 11 is probably more useful to state administrators than it is to local personnel. A detailed simulation of the effect of input and process on product might be more useful to local decision makers.

Figure 11 depicts the interactions for the school sub-system with the World System. Within the World System are several sub-systems (or impacts) represented by a rate and level unit. Take, for instance, the "Employment of Completors" sub-system (58). The "Drop-out Employment Multiplier" represents the effect of: a) vocational school drop-outs; and b) those whose applications were rejected by the school sub-system and subsequently dropped out. Then there are also the effects of the number completors (Completors on Employment Multiplier, (72)), and the reputation of past vocational graduates. Note that there is an indirect feedback loop via skill demand between present and past vocational school graduates. Hence, we have the possibility of the school system flooding the market for certain skills. Note also that skill demand is a fairly aggregated function which one might want to further break down in a more detailed model.

Indirect feedback loops can be traced back from the World System to the school sub-system via the Application Rate (50). The application rate is a function of previous admission policies (Assignments on Applications Multiplier (49)), Social-Economic Mobility (Distribution on Applications Multiplier

(51)), and the demand for the skills taught by the schools (Job Demand or Application Multiplier (53)). The simulation would, therefore, contain procedures for modifying Applications Rate (50) whenever there was a change in the levels of any of the sub-systems (impacts) controlling the multipliers which impinge upon this Rate. There is a second feedback from employment levels via income distribution on tax revenue. Note that there is a box for every impact (although the reverse is not true) for which we have an equation. Note, for instance, that the interactions of Skill Demand (74) on several other impact sub-systems is reflected by its presence in the corresponding equations of Figure 10. Hence, changing the parameters of Skill Demand would require us to recompute each of those impact values. This would have to be specified in procedure form. Finally, note that the diagram is incomplete. Many relationships, particularly within the school system, are merely hinted at. And, we have not fully described even one functional relationship or much of the procedure necessary to make this simulation operational. The purpose of this example is to illustrate some of the techniques and modeling requirements of simulation methodology.

The results of each simulation run (there is one run for each set of starting levels) would be displayed in the form of a set of curves (one curve per impact) projecting outcomes into the past as well as into the future. A very hypothetical example of these curves is shown in Figure 16. Let us assume that the occupational education sub-system is the sole source of skilled workers and that a user has asked what would happen if we continue current expenditures (adjusted for inflation, increased maintenance with time, etc.)

The curves show that with increased capital investment, the demand for skilled labor increases and the demand for unskilled labor (as exemplified by drop-outs) decreases. The number of skilled workers is increasing linearly. That is, the occupational education sub-system is adding no more than a fixed number (the number of seats available in senior classes) of skilled workers to the labor. On the other hand, in a growing economy, demand for skilled labor increases exponentially. The result is a shortage of skilled labor which prevents a technologically advanced economy like ours from expanding. It does not pay to invest further in sophisticated production processes unless there is a simultaneous increase in the number of skilled workers to man the equipment. The competitive position of the state economy will eventually decline vis-à-vis the rest of the nation.

During the initial phase of decline the better established firms are able to delay the day of reckoning. They are able to absorb the small increase in skilled labor that is being turned out as well as most of those who are laid off by the weaker companies. Hence, the declining phase of the economy has to progress to a certain point before the demand for skilled labor declines. Eventually, these larger employers begin to pull out and this produces a rapid decline in the demand for skilled labor.

Meanwhile, the increasing population coupled with the decline in white collar positions results in an increase in the number of vocational school applicants. This increase in applicants occurs in spite of the counter influence of the decline in the number of blue collar positions. Of course, the number of students turned away increases and this increases the drop-out rate. This exponentially increasing drop-out rate results in an exponentially decreasing employment rate reflecting the ever lower demand for unskilled labor.

Decreasing investment and employment results in a decline in tax revenues. Maintaining occupational education services at a fixed level in the face of declining revenues reduces the funds available for other social services. The user would conclude that continuing the present level services would exacerbate current social problems. (Again, assuming that the occupational education sub-system is the sole source of skilled workers). He would then proceed to test the effect of a range of alternate policies. Using the latter results, an informed decision could be made on whether or not a feasible vocational education expansion program would significantly countermove these trends.

Figure 12 depicts the stages of development of a simulation model. One thing is clear, this is a very dynamic process. It encourages one to think of alternatives, i.e., alternate outcomes and alternate models. Working with a simulation is a highly interactive process and this process requires a highly interactive environment. Special simulation packages have been developed which reduce the effort required to formulate and reformulate the model. And, once the model is formulated, these packages provide easy-to-use data entry, parameter generation and results analysis services. These benefits are provided in the form of a language with its own grammar and vocabulary specifically designed to help one define the problem and then formulate alternative models. However, something more is required if the computer system is to be truly interactive and this is described in the next section on system development.

IV. SYSTEM DEVELOPMENT

Two phases of system development can be anticipated. Phase I would involve decisions: a) on the kinds of pre-determined analyses that would be provided; b) what data is needed; and c) what pre-determined kinds of predictions we would like to be able to make. Having made these decisions, the data base would be designed. Analytical and predictional programs would be written and storage space for them and their products would be allocated. Finally, system programs to handle data entry, data and program base mapping, and requests for analyses would be written. At the end of Phase I, a pre-determined set of services would be available, subject, of course, to modification by the system's manager. Analytical programs will be designed to use the data base efficiently. And analyses will provide sub-results for each other whenever possible.

The data and program base maps (generated by the data accounting function) will be in the form of an index to the system's contents. Included in this index will be; a) data and descriptives; b) analyses and optional variations on analyses; c) sub-results; and d) forecasting services. Supplemented by appropriate text, this index will be made available to users in report form. Requests for index items will be filled on a batched processing basis at the system site.

Phase 2 would feature initially on-line remote entry of data and requests. Set up at several regional offices would be visual display scopes with light-pens and keyboards for conversing with the system and a teleprinter for hard copy. This system would require the development of an elementary conversational language and time-sharing capability. Later, the capabilities of the conversational language would be expanded to allow users to request temporary modifications, additions, etc., of the "pre-determined" sub-systems II and III analyses.

During both phases, we will make use of the suggestions of members of the user community. Continued contact with users will be maintained and will influence modifications, additions and other improvements in a system services. Reports filling users' requests will contain suggestions on how users may make better use of them. The analyses will be designed to suggest alternatives rather than single so-called "optimum" decisions.

It is not the function of the system to direct the choices of the decision maker. Rather, the system is designed to broaden the decision maker's perspective and to increase the quality and the relevance of the information available to him. Users will be encouraged to evaluate and interact with the system in all phases of development. The quality of this interaction will take a quantal jump when we go on-line in Phase II, enabling users to converse directly with the system and to see the results of the alternatives they have chosen without delay.

A. INPUT

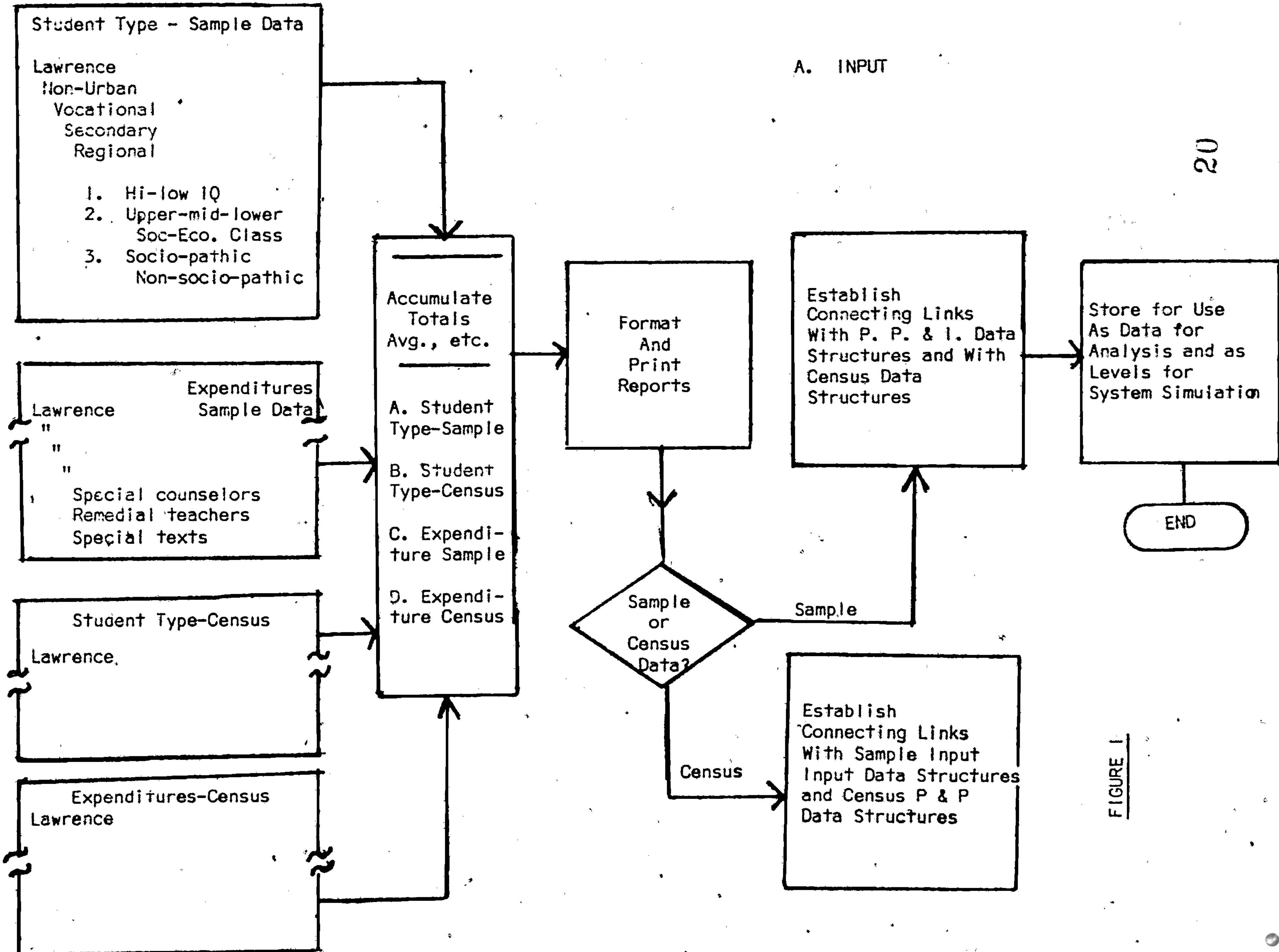


FIGURE I

A. Input Data File: Address = 1

DATA	100	1000	10000

B. Process Data File: Address = 100

DATA	1	1000	10000

C. Product Data File: Address = 1000

DATA	1	100	10000

D. Impact Data File: Address = 10000

DATA	1	100	1000

Figure 2: Example of Connecting Links in Data-Entry Storage Files

B. PROCESS

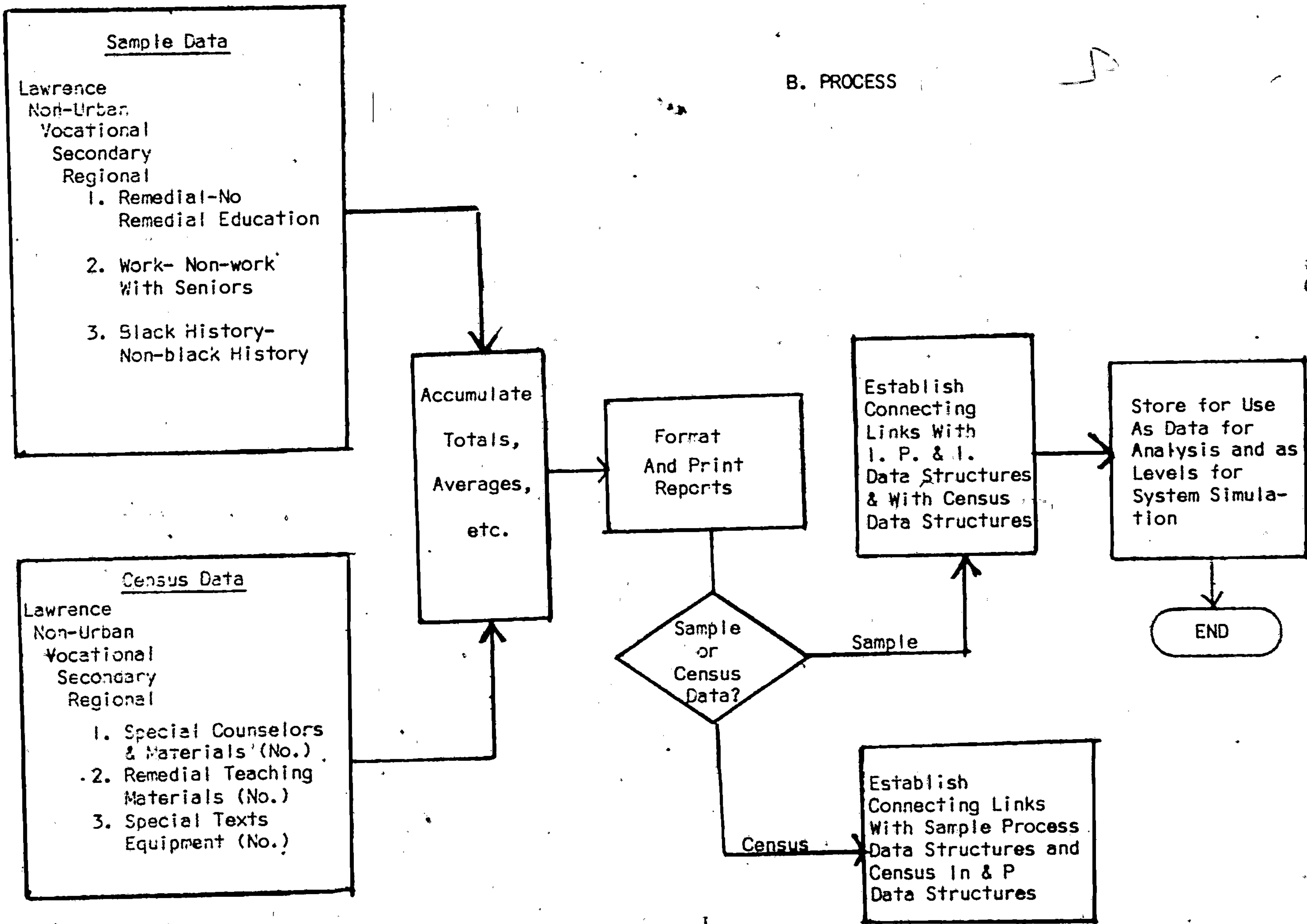


FIGURE 3

C. PRODUCT

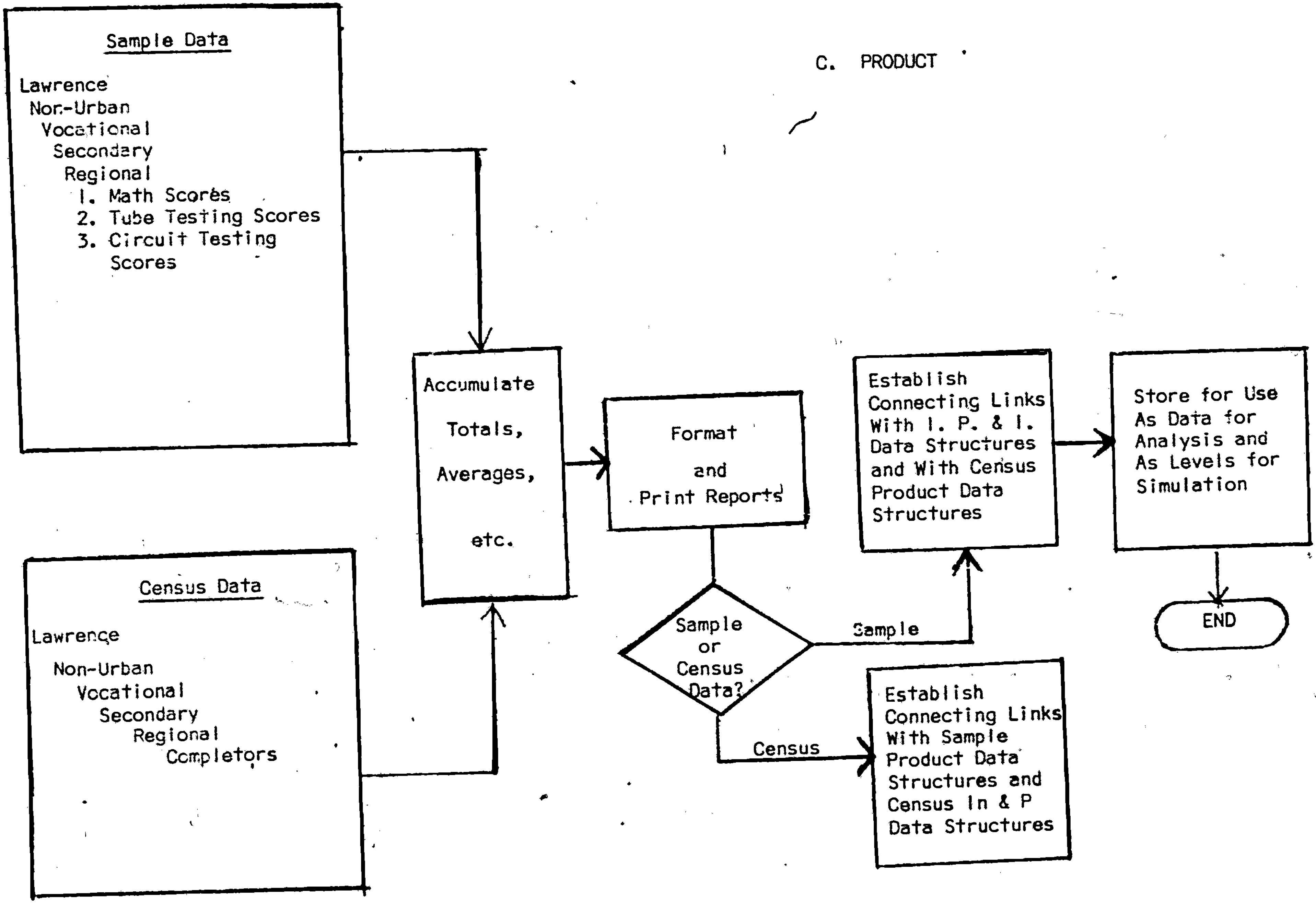


FIGURE 4

D. IMPACT

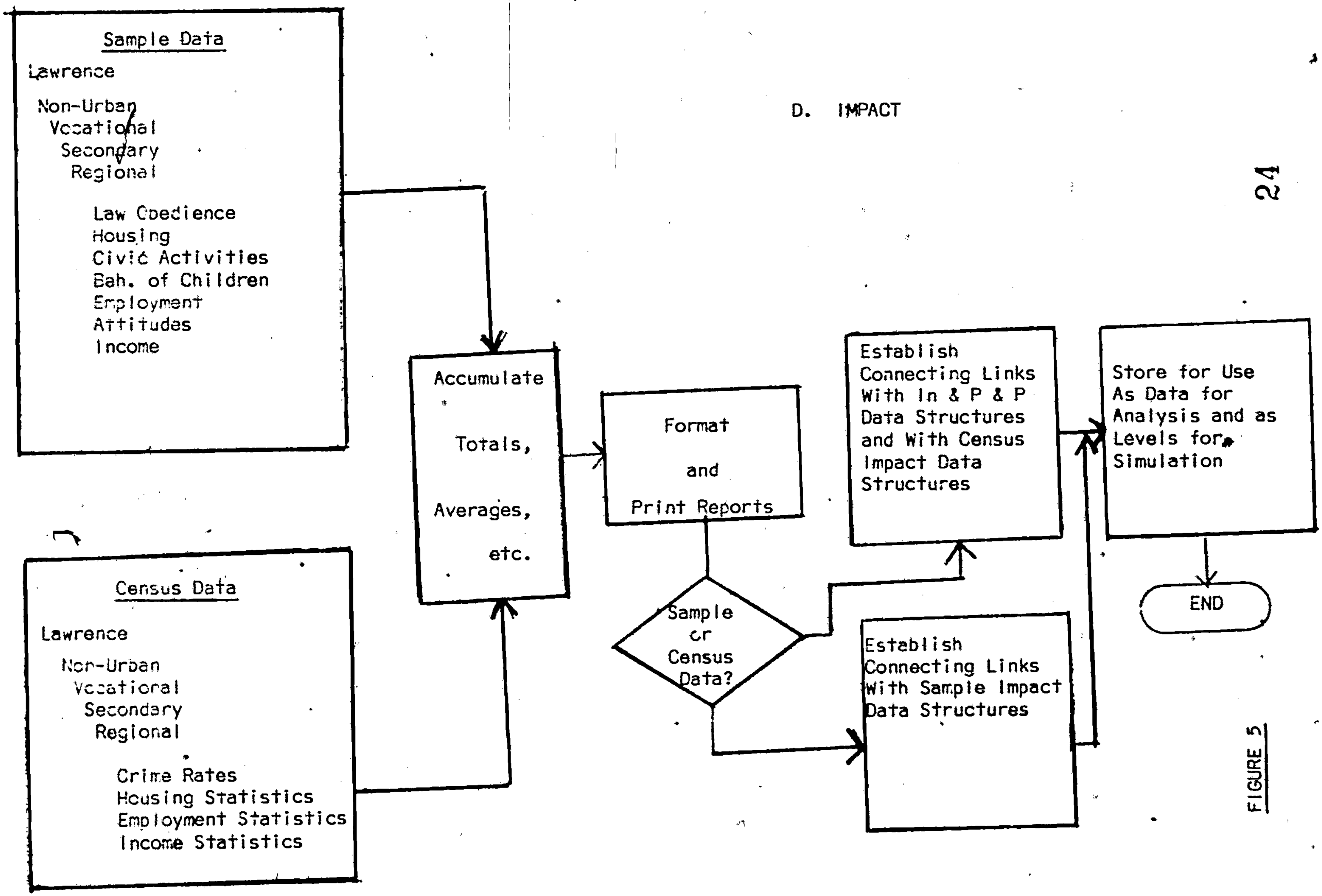


FIGURE 5

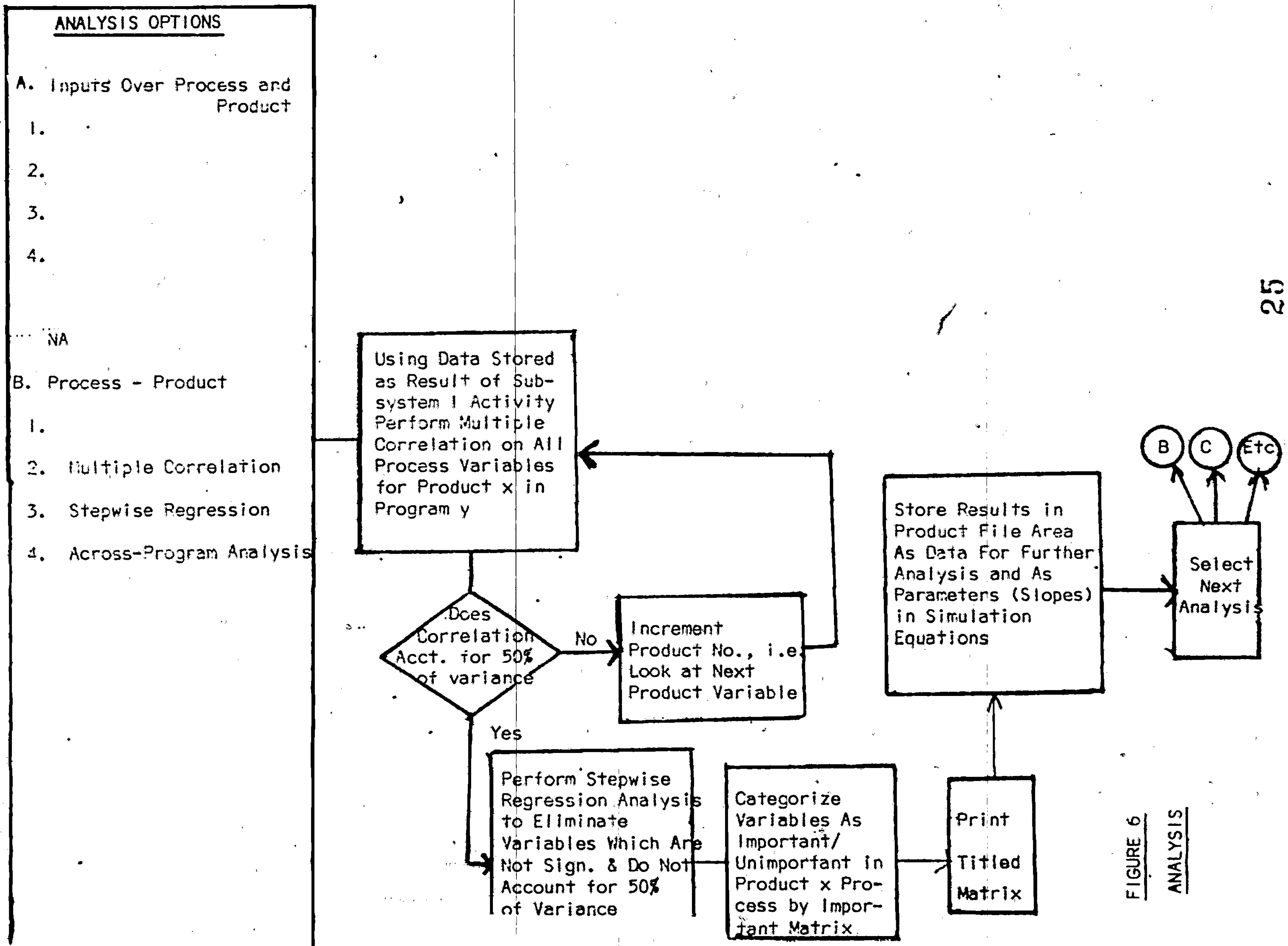


FIGURE 6

ANALYSIS

C. Product - Cost

1. Costing B Variables

2. Costing Input Data

3. Sensitivity Analysis

4. Costing Product

5. Historical Comparisons

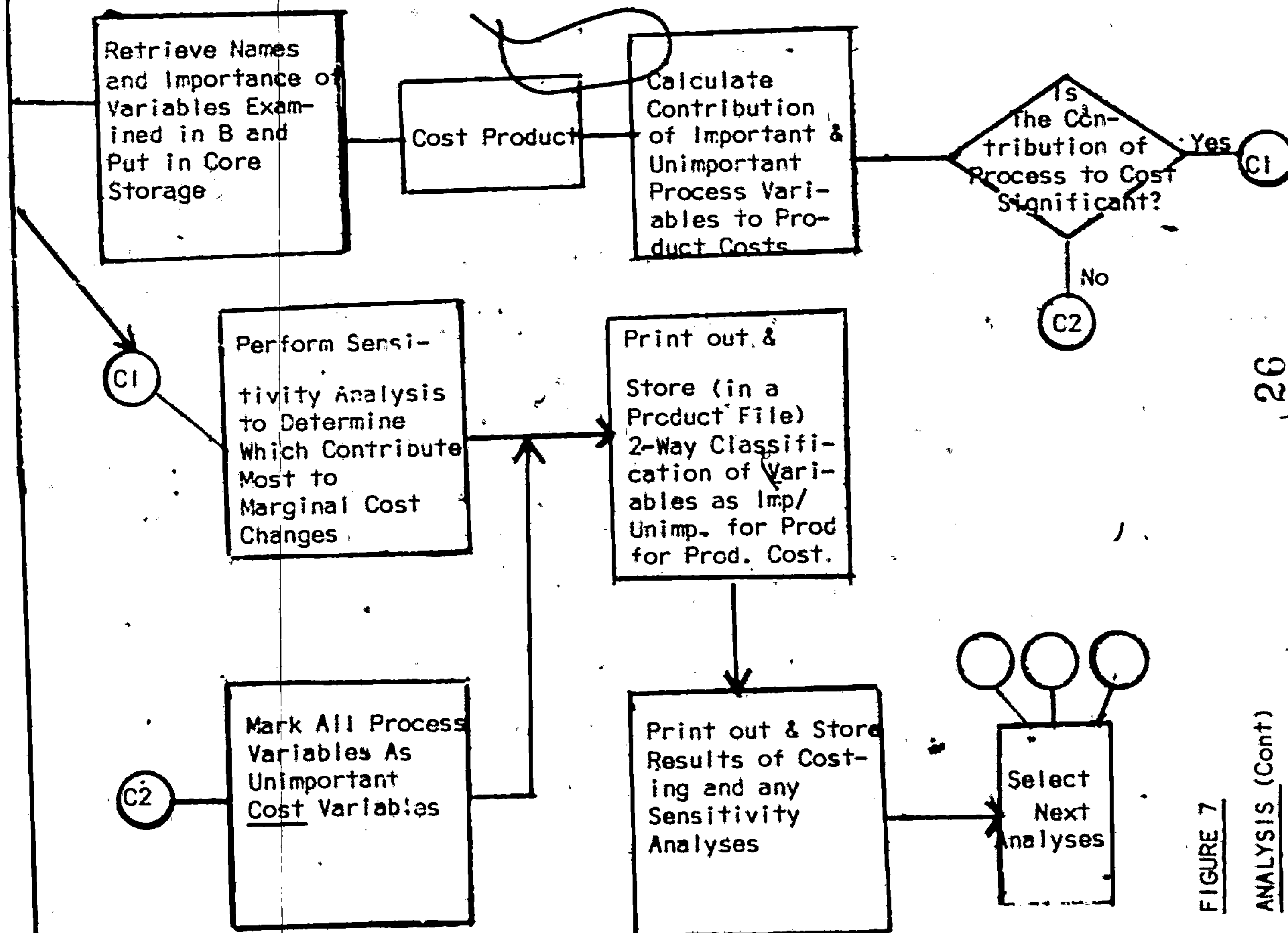


FIGURE 7

ANALYSIS (Cont)

D. Product - Impact

1. Employment
2. Welfare Costs
3. Voting Behavior
4. Drop-out Rates
5. Socialization
6. Income Redistribution
7. Children's Academic Performance

E. Process - Impact

1. Employment
2. Welfare Costs
3. Voting Behavior
4. Drop-out Rates

F. Impact Cost

1. Employment
2. Welfare Costs
3. Voting Behavior
4. Drop-out Rates

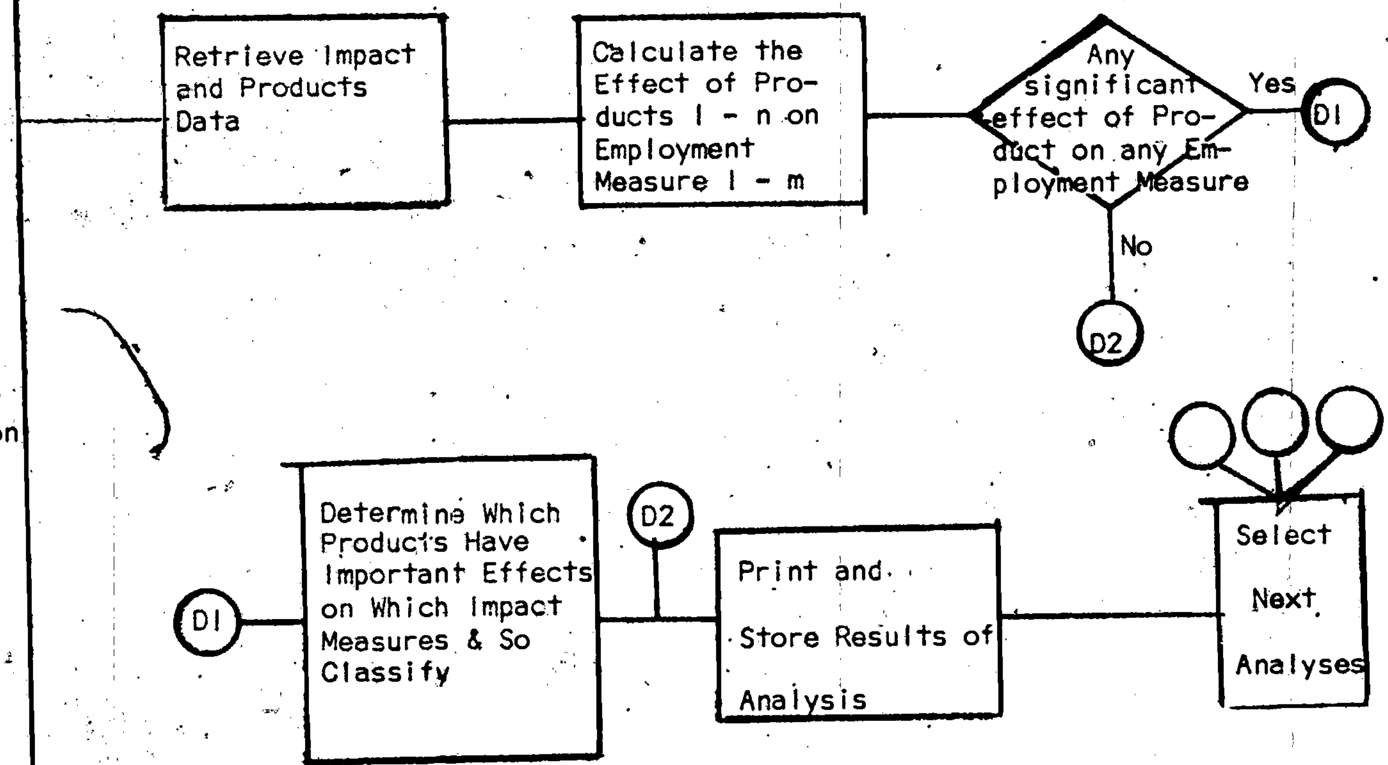


FIGURE 8

ANALYSIS (Cont.)

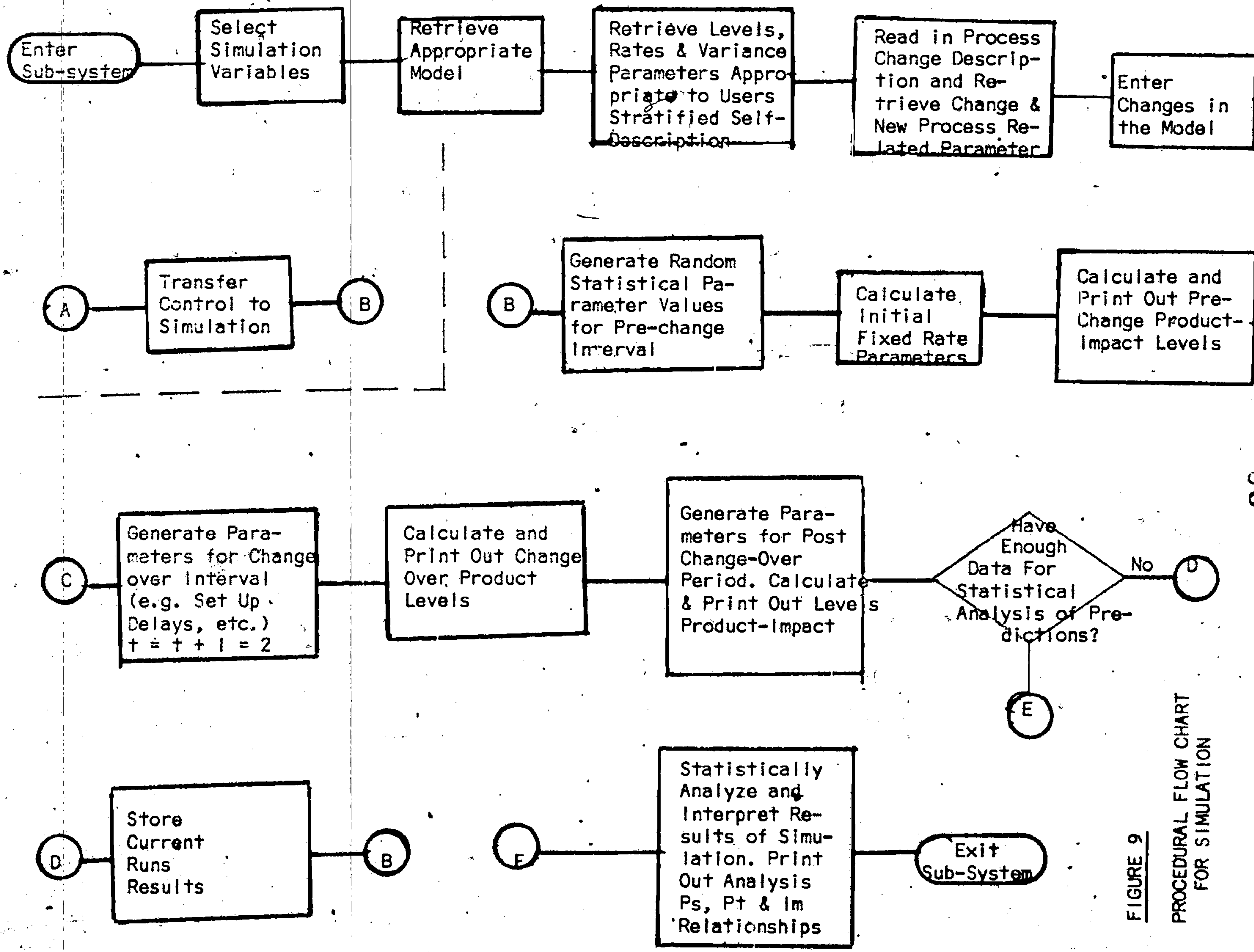


FIGURE 9

PROCEDURAL FLOW CHART FOR SIMULATION

Figure 10: Simulation Variables

IM_1 = Number of Drop-outs

IM_2 = Number of Drop-outs Employed

IM_3 = Income Distribution

IM_4 = Number of Current Vocational School Completors Employed

IM_5 = Number of Jobs Per Skill

IM_6 = Number of Past Vocational School Completors Employed

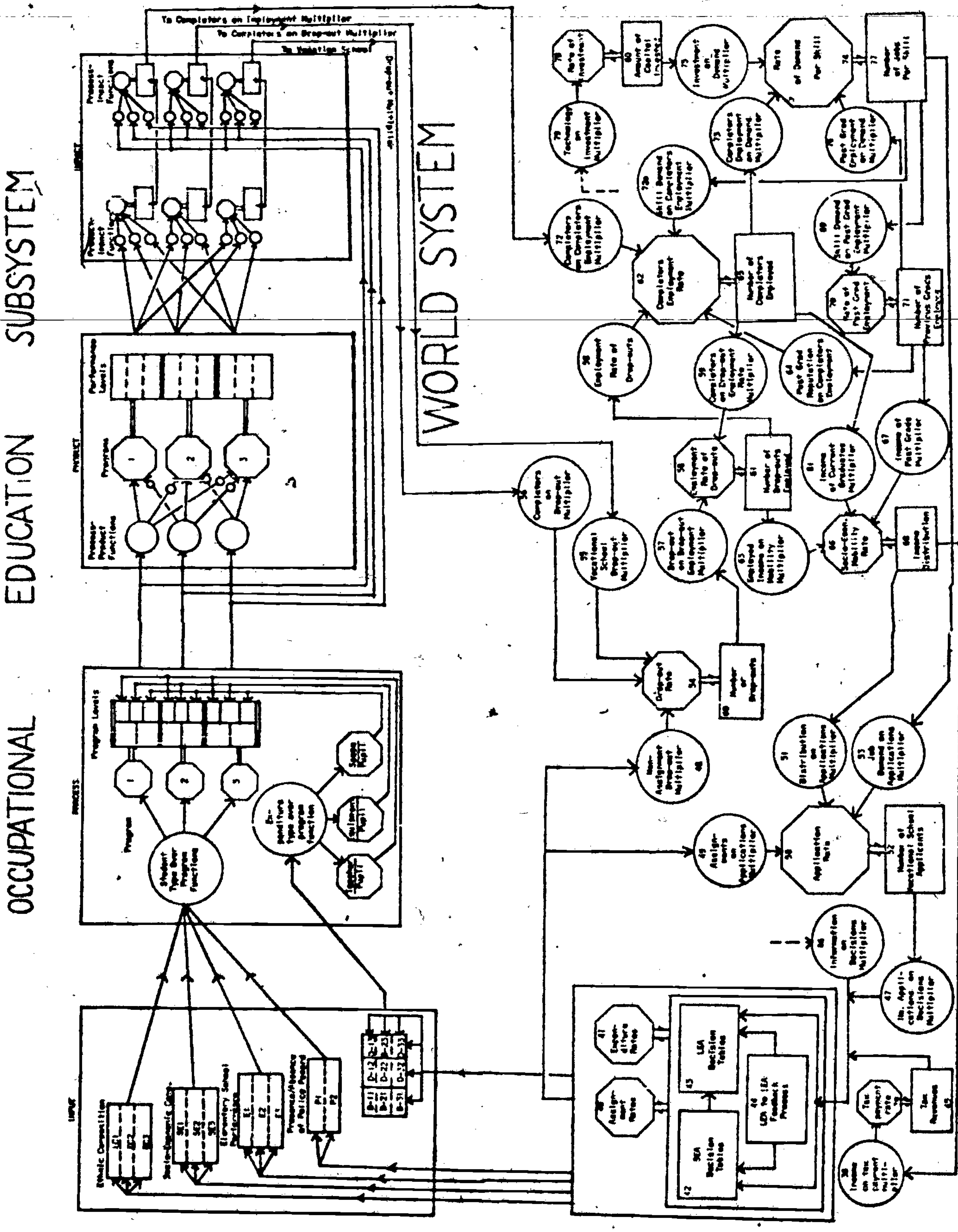
IM_7 = Applicants to Vocational School

IM_8 = Administrative Decisions

1. $IM_1 = IM_1 + (IM_1).F_1$ (Completors, Rejections, other sources).
2. $IM_2 = IM_2 + (IM_2).F_2(IM_1, IM_4,$ other sources).
3. $IM_3 = IM_3 + (IM_3).F_3(IM_2, IM_4, IM_6,$ other sources).
4. $IM_4 = IM_4 + (IM_4).F_4(IM_2, IM_5, IM_6,$ Completors, other sources).
5. $IM_5 = IM_5 + (IM_5).F5($Capital Investment, $IM_4, IM_6,$ other sources).
6. $IM_6 = IM_6 + (IM_6).F_6(IM_5,$ other sources).
7. $IM_7 = IM_7 + (IM_7).F_7(IM_3, IM_5, IM_8,$ other sources).
8. $IM_8 = IM_8 + (IM_8).F_8(IM_7,$ Information, Revenue, other sources).

FIGURE II

DIAGRAM OF SIMULATION MODEL RELATIONSHIPS



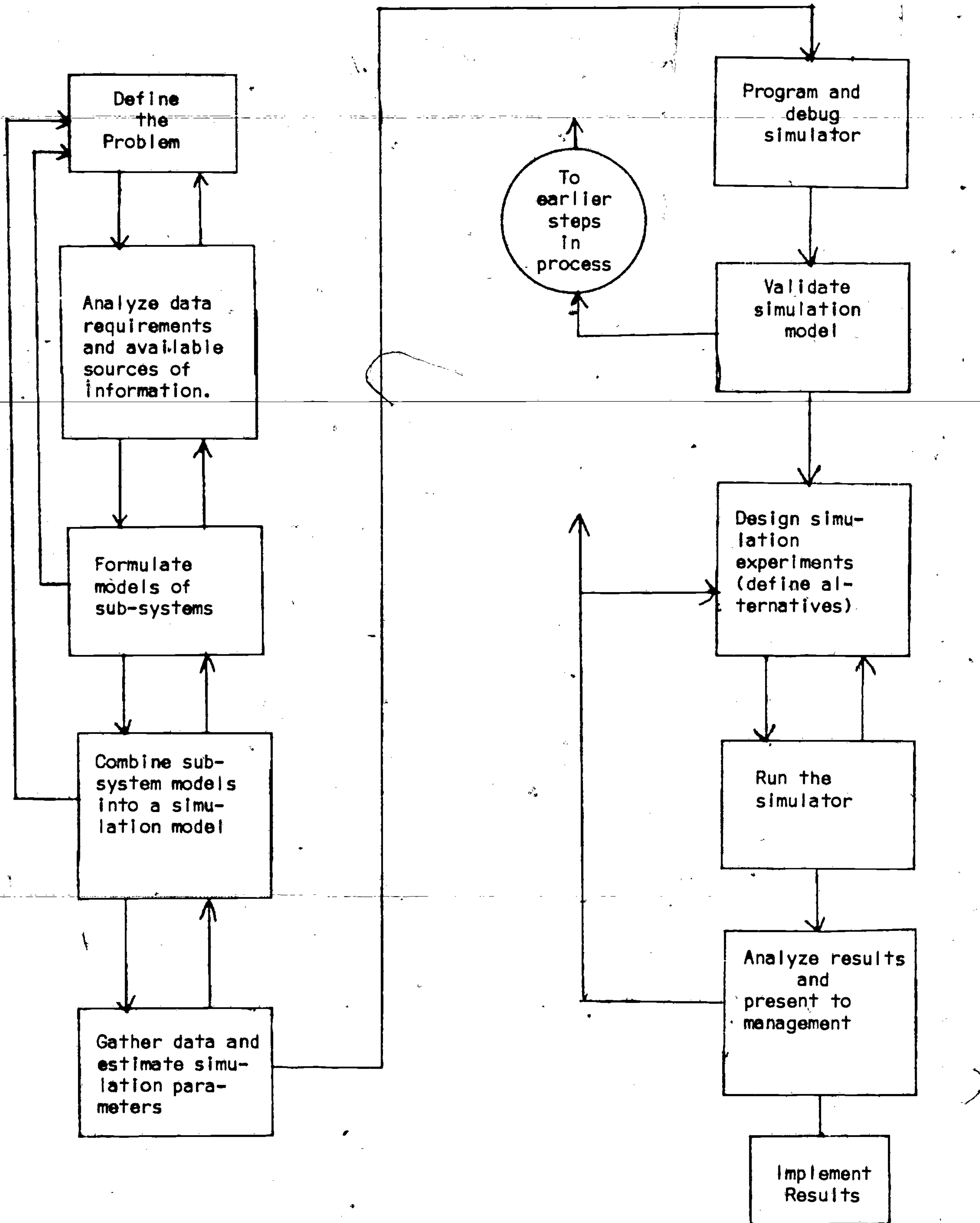


Figure 12: Steps in Simulation Development (From Emshoff & Sisson)

SIMULATIONS - COMMON FEATURES

- Create Random Numbers
- Create Random Variates
  - distribution -
- Advance Time, either by One Unit or to Next Event
- Record Data for Output
- Perform Statistical Analyses on Recorded Data
- Arrange Outputs in Specified Formats
- Detect and Report Logical Inconsistencies and Other Error Conditions

SIMULATIONS OF DISCRETE PROCESSES

- Determine Type of Event
- Call Sub-routines to Adjust the State Variables as a Result of the Event
- Identify Specific State Conditions
- Carry Out Procedures Contingent Upon the Identified State

Figure 13: Common Simulation Processes (From Emshoff and Sisson)

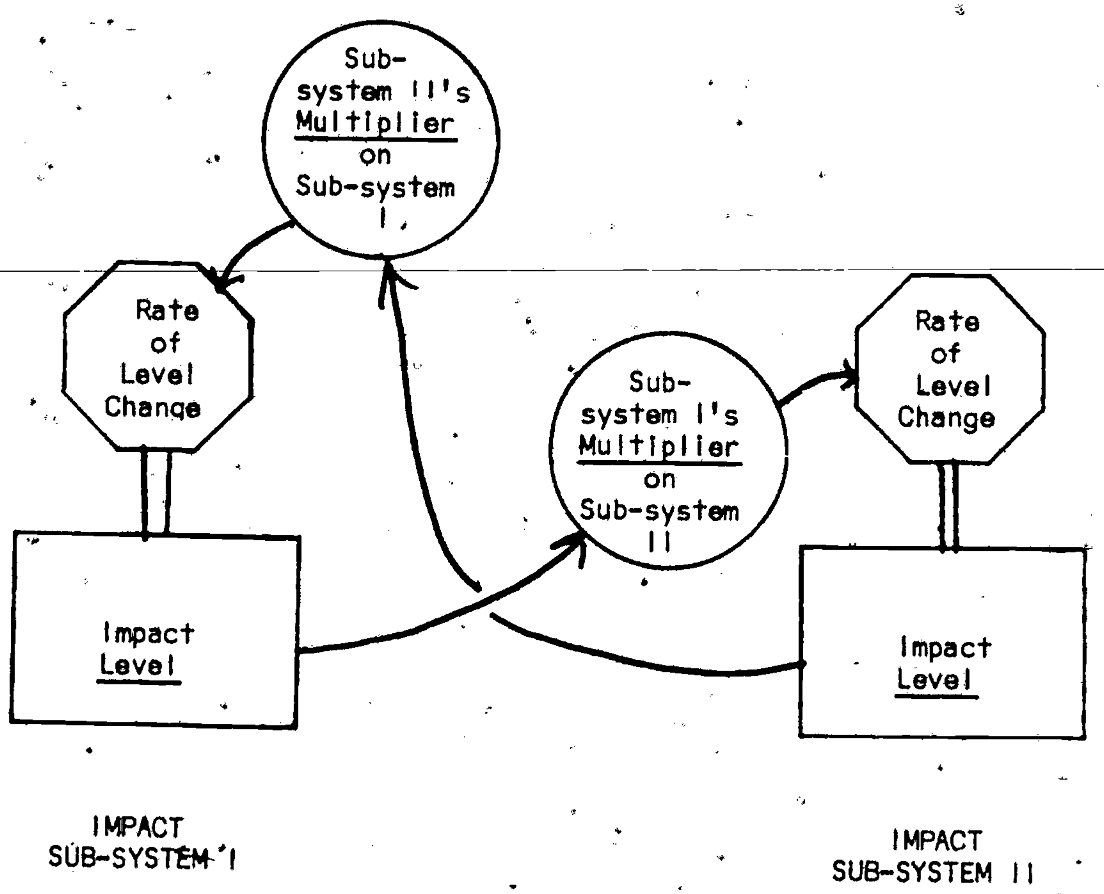


Figure 14: A World Dynamics Feedback Loop

		Contributions of Process Variables to Product Scores: Important	Contributions of Process Variables to Product Scores: Unimportant
Contribution of Process Variables to Product Cost	Important	Variables A . B . C . D . E . F . G	Variables M . N . O . P . Q
	Unimportant	Variables H . I . J . K . L	Variables R . S . T . U . V . W . X . Y . Z

Figure 15: An Example of Combining Analysis Types

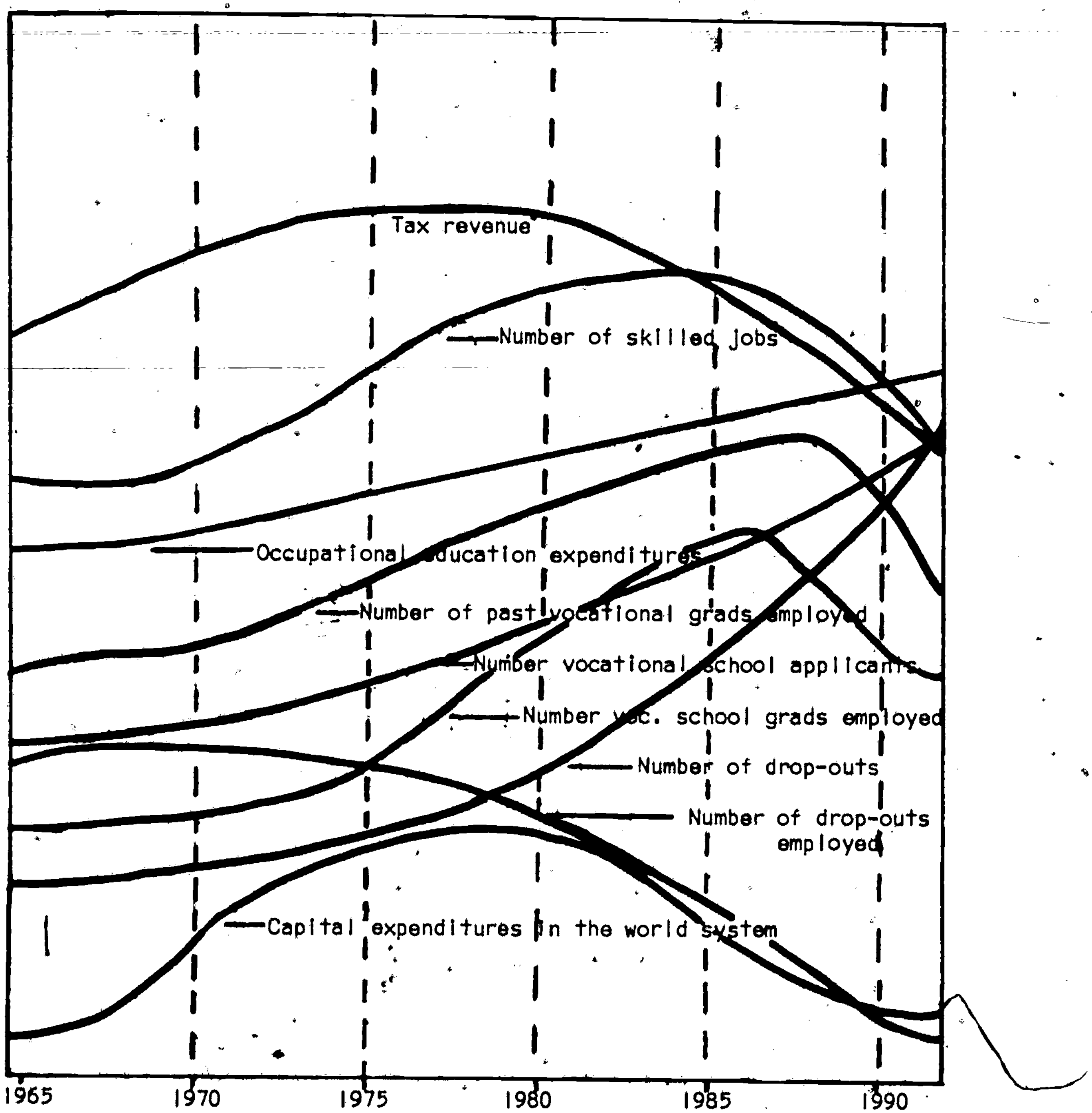


FIGURE 16

Projected Impact of Holding Adjusted Expenditures Constant