1-1-1971

A validation study of a computer simulation model for an individualized curriculum.

William Webster Foley
University of Massachusetts Amherst

A VALIDATION STUDY OF A COMPUTER SIMULATION MODEL FOR AN INDIVIDUALIZED CURRICULUM

A Dissertation Presented

By

WILLIAM WEBSTER FOLEY

Submitted to the Graduate School
of the University of Massachusetts
in partial fulfillment of the requirements
for the degree of

DOCTOR OF EDUCATION

May 1971

Major Subject: Administration

A VALIDATION STUDY OF A COMPUTER
SIMULATION MODEL FOR AN
INDIVIDUALIZED CURRICULUM

A Dissertation

By

WILLIAM WEBSTER FOLEY

Approved as to style and content by:

(Chairman of Committee)

(Head of Department)

(Member)

(Member)

Kenneth H. Blanchard

(Member)

MAY 1971

ACKNOWLEDGEMENTS

The author would like to express his thanks and appreciation to the following people for their contribution to the completion of this thesis.

To Dr. G. Ernest Anderson, Jr., for his moral support and technical advice in the preparation of this thesis.

To Dr. William C. Wolf, Jr. and Dr. Kenneth H. Blanchard for their advice, support, and careful reading of the manuscript.

To Dr. Albert S. Anthony for his careful reading of the manuscript.

To Frederick A. deFriesse for his assistance and advice in computer programming and data extraction.

To Dr. Richard Ferguson and the staff of the Learning Research and Development Center in Pittsburgh, Pennsylvania, for making the data available and for the valuable feedback received in regard to the value and use of simulation models.

And, finally, to my wife, Gladys, without whose patience, love, and understanding, this project could never have been attempted and surely never completed.

TABLE OF CONTENTS

LIST OF TABLES

Table | Page

LIST OF ILLUSTRATIONS

CHAPTER I

INTRODUCTION

The use of simulation in educational analysis in contrast to the relatively long history of its application in engineering and scientific studies is a recent development. Teichroew (1965) states that simulation methods in engineering and scientific studies have been in use well over fifty years, substantially predating its use in business, economics, and education. The technique emerged as a major tool of analysis in business and economics in the mid 1950's, as digital computers started to become widely available; most of the literature in the area has been published since 1955 (Meier, 1969). In education, interest in simulation techniques as a possible instructional device began in the late 1950's; interest in it as an analytic tool began somewhat later. There are very few references listed for simulation in the Education Index prior to 1963. Simulation by means of digital computers is now beginning to find rather wide acceptance in the analysis of business administrative and economic problems, and as a result it is being considered more widely as an acceptable tool in educational planning, evaluation, and management (Meier, 1969).

It is becoming increasingly difficult to discuss simulation in educational circles. There appears to be little uniformity of terminology, which is to be expected in such an emergent field; and such statements as "Simulation, as we discuss it here, will mean many things to many people" or "Simulation is becoming a very popular but difficult to define technique" are being encountered more and more frequently in educational seminars and journals. Several examples will demonstrate the broad meanings that have been attached to simulation. In social science, simulation may be either the constructing and manipulating of an operating model, or it may be a representation of reality (Guetzkow, 1962). For other social scientists it is the study of structures in a lab (Zelditch and Evan, 1965). In education simulation may mean the creating of games (Cruickshank, 1966) or a model of a system (Beaird and Standish, 1964; Cogswell, 1965). In business it could refer to a decision-making exercise structured around a model of a business operation (Greenlaw, Harron, and Rawdon, 1962). In many cases, it is difficult to be entirely sure of what the author means unless the what, how, and purpose of the simulation is studied in detail.

McCormich (1964) states, "Simulation consists of some type of reproduction or representation of an actual or conceptual physical object, system, process, or situation, or of a theoretical construct." Twelker (1969) states, "It should also be emphasized that the fact that one person thinks of simulation as a device while another person thinks of simulation as a technique for setting up the device, is not undesirable. It simply points up the two acceptable meanings of the word. The usuage which might be adapted by an individual would depend largely upon his discipline, and the use to which the simulation is put."

In this study simulation will be defined according to Mogenthaler's (1961) definition, "to duplicate the essence of the system or activity without actually attaining reality itself." Therefore, simulation as used in this study will be the use of a model to represent over time essential characteristics of the system or process under study. The dynamics of the behavior of the system represented may be inferred by the operation of the model.

Statement of the Problem

Validation of models and interpretation of experimental results present practical and theoretical questions in the use of simulation techniques that are not completely resolved. How is the user to know that a model represents the process under study? How is the user to test the results to yield predictions that are empirically accurate? At times, this question of validation becomes a question of credibility; i.e., what evidence of the model's ability to describe and predict aspects of the system is required before a user will utilize the model.

This question of validity arises in every simulation study. How close do simulated and real events have to match in order to have a valid model? How well does the model describe a portion of the real world? How well can the model predict something that verifyably happens in the real world based on parameters determined from the same population? How sensitive is the model to relatively minor, random, and uncontrollable changes in system parameters? What improvements in the model should be made as a result of experience with real data? These questions are questions that should receive attention in a validation study.

There is no question that simulation models can be created for almost any situation in which one can precisely state relationships among variables. However, the usefulness of the model could rest on its validity or on the criteria used to judge its validity. At present little work has been done to answer questions concerning validation of models or to describe procedures for conducting validation studies; the emphasis has been on model construction. The current literature concerning validation of models is reviewed in Chapter II of this study.

Objectives of the Study

The primary objective of this study is to use data from an operating educational program in order to validate the basic assumptions underlying EDSIM I, a simulation model that traces the interaction of three probabilistic elements in an educational program and produces synthetic output regarding this interaction in terms of student time in the system. The validation will be done from the viewpoint of a potential user in order to increase the confidence an educational administrator might have in the technical coding of the model, its ability to describe the system under study, and its ability to predict student time in the system. The secondary objective is to examine the relationships between variables in the model and student background data to determine possible predictors that might be used to improve the model.

Basic Definitions Used in This Study

Model

METEP: Model Elementary Teacher Education Program

PROFILE: A major division of the METEP curriculum. It designates the number of performance criteria that need to be mastered in a specific curriculum area (Human Relations, Behavioral, *etc.*)

PERFORMANCE CRITERIA: A specific skill that must be mastered.

INSTRUCTIONAL ALTERNATIVE: An instructional event (a reading, a lecture, *etc.*) that may be used in mastering a specific performance criteria.

System

IPI: Individually Prescribed Instruction

LEVEL: A major Division of the IPI Mathematics Program. Eight levels of work difficulty comprise the mathematics component (Levels A through H).

AREA: One of the thirteen subject matter topics offered in a given level (numeration, place value, addition, *etc.*)

UNIT: The group of skills that comprise the area of a given level. A unit is specific for one area in a level. In Table 3.1, a unit is the cell at the intersection of an area and level.

Specific Objectives

I. To list and describe the variables and parameters used in EDSIM I for tracing the interaction of elements in regard to time spent by students in an educational program that is designed for individual progress and based on a curriculum that defines performance in relation to a series of behavioral objectives organized into areas and levels.

II. To describe the mathematics component of the Individually Prescribed Instruction program as it is currently in operation at Oakleaf Elementary School in Baldwin, Pennsylvania, in terms of the components of EDSIM I.

III. To use data available from the mathematics component of the program to analyze the following basic assumptions underlying the model in regard to time spent by students in a system designed for individual progress.

Basic Assumptions to be Analyzed:

1. The probability of pre-testing out of a unit if taken as a function of curriculum area and level will reflect the system operation in regard to the initial number of units a student must participate in to complete a specific level.
2. The probability of achieving a satisfactory performance on the unit post test if taken as a function of curriculum area and level will reflect the system operation in regard to whether a student must do additional work in a specific level.

3. Time spent to accomplish adequate performance in a unit may be determined by randomly sampling one distribution of service times based on student time in a unit both before and after a post test.

IV. To improve the EDSIM I model so that it will more accurately describe the interaction of student time elements in the system.

V. To use the improved model to predict student time to completion for selected samples of students in a specific level for the 1969-1970 term using input data from 1967-1968 and 1968-1969 school terms.

VI. To examine student attributes as a function of performance in the system so that recommendations can be made concerning future construction of models that predict student time in an educational program.

Procedures

After data tapes containing data from the IPI project were edited for missing and incomplete data, pertinent data was extracted and detailed accounts given of the assumptions used in the extraction of the data. These data were then coded into pre-test and post-test probabilities and service time distributions so that they could be used as input for the EDSIM I simulator. Runs were made to test the computer programming for technical flaws. After the model was determined to be functioning correctly, runs were made to evaluate the basic assumptions underlying the model, and minor changes were made to improve the simulator so that it better described the program under study. Experimental runs were made to test the predictive ability of the model and to investigate the usefulness of the technique in regard to furnishing worthwhile information to educators. Correlations were made of selected background data and time data to examine performance as a function of student attributes in a program designed for individual progress.

Validation Criteria

Synthetic output produced from runs of the model are graphically displayed with empirical data from the educational program so that selected comparisons may be made in regard to (1) frequency distributions of time in a level; (2) range of time for completion of a level; (3) variability of time to completion; (4) highest time required in a level; (5) lowest time required in a level; (6) average time in hours spent in each area of a specific level. Where appropriate the Kolmogorov-Smirnoff Two Sample Test was used to test how well the simulated distributions fit the real distributions observed in the data from the educational program.

Significance of the Study

Conway (1961) states that the assertion that management systems can be simulated is a statement of faith rather than fact. The question of the usefulness of computer based simulation models as a planning, evaluating, or decision-making tool for educational administrators is surely one that is under debate at the present time. Many of the models that are being built by technicians to simulate educational systems are being placed on a shelf waiting for adequate information systems to be developed so that pertinent data about the system can be collected in order to test the validity of the assumptions underlying the model. It becomes a question of whether simulation models are tools to be used by educators or analytic exercises for development of model builders. Since all the EDSIM models developed at the University of Massachusetts use the same basic assumptions in predicting student time in the system, this study will contribute to the confidence a user might have in the synthetic output produced from the student time phase of these models. This study should also contribute information concerning techniques of data management, data interpretation, and procedures that are pertinent to the conducting of experiments with simulation models. Meier (1969) states, "In contrast to analytic models that yield solutions to problems, simulation experiments yield results that must be treated as experimental data. Development and trials of experimental procedures is an area often given insufficient attention."

CHAPTER II

RELATED LITERATURE

This chapter shall consist of a review of pertinent literature in regard to simulation. It becomes necessary if a user is to understand how to use a simulator effectively or is to judge the validity of its results, that he be aware of the purposes and uses of simulation models as analytic tools, as well as approaches that might be used in validating a model. This chapter shall be divided into three parts so that literature dealing with purposes, uses, and validation of models may be reported.

Purposes

Simulation was a natural development in the discipline of operations research. Operations research evolved out of the efforts to provide formal, efficient decision-making techniques for the design of air defense operations in Britain in World War II and progressed by improving middle management decisions (Emshoff and Sisson, 1970). As techniques for solutions to middle management decisions were developed and utilized, operations research turned its attention to higher level management decisions and found that a large part of these decisions were qualitative in nature. It was generally found, however, that these qualitative decisions could be improved if appropriate quantitative aids were available. This need for quantitative techniques as an aid for what had been total intuitive decisions, combined with the availability of large scale electronic computers has led to the development and growth of computer based simulation models. (Emshoff and Sisson, 1970)

The need and possible basic purpose of simulation can be stated very well by a quote from one of the more recent books in the field.

> "Managers who are living in a world of rapid change and extensive interaction, must continually improve their own decision-making skills or end up reacting to crisis rather than controlling activity. Apprenticeships and experience are not enough: today judgement and intuition are barely put into use before change occurs. A formal and efficient technique is needed to augment the manager's experience. The technique must be formal--that is, capable of precise documentation--so that it can be learned quickly and applied to new situations. The technique must be efficient so that its cost does not increase in proportion to the complexity of the situation. . . computer simulation is a technique that will fulfill these needs." (Emshoff and Sisson, 1970, p. 2)

Rowe (1968) reports that computer simulation can be considered to be an attempt to model the behavior of a system in order to study its reaction to specific changes. It is seldom an exact analogue of an actual system but is usually an approximation of time dependent activities within the system. If one can accurately define the properties and elements of the system then it is possible to study the behavior of the system by tracing through the computer the simultaneous interaction of key variables. Such a model of a system will usually indicate relationships which are otherwise not obvious and should have the capability of predicting system behavior. (Rowe, 1968) Rowe further states, "models are merely the basis for testing new ideas and should not become ends in themselves."

Geisler (1969) states that a user who wants a closed solution to a formulated problem should use a linear programming model or some type of sequential decision theory technique. It is his contention that simulation should be considered more of a heuristic process in which the user attempts to reach an optimal solution by iteration of the model. There is an increasing awareness that educational as well as business and economic problems should be examined in terms of the total system. Since iteration of the model will allow for observations to be made in regard to interactions between parts of the system, simulation is a tool that offers an excellent opportunity for adopting a "systems" approach. (Geisler, 1969)

Meier (1969) reports that simulation offers a unique opportunity to observe the dynamic behavior of complex interactive systems. He states that a carefully constructed realistic simulation model provides a laboratory environment in which to make observations under controlled conditions. It can provide an experimental environment for testing hypotheses, decision rules, and alternative systems of operation under a variety of assumed conditions. When there are no practical analytical approaches available and when it is too costly or impossible to experiment with the actual system under study, simulation is the most recommended technique and the one most likely to be successful. (Meier, 1969)

There are, of course, many processes, such as the decision process in an organization or the interaction of a system with its environment that are too complex to be represented by more formal mathematical structures. Sometimes even those processes that can be formulated mathematically such as large multichannel, multistage queuing problems and complex scheduling problems may defy attempts at solutions by direct mathematical procedures now available. In such cases, simulation may prove to be the only practical method of analysis. (Meier, 1969)

Simulation is sometimes thought of as a technique to be used after all other approaches have been examined and found wanting. Meier (1969), however, points out several unique characteristics of simulation that could justify its use in preference to other techniques. He lists the following characteristics:

1. Simulation permits direct and complete observation of the dynamic behavior of processes.

2. Time can be compressed or expanded in a run to provide observations in any desired degree of detail.

3. Each run of a model may be viewed as an experiment from which observations of a variety of system characteristics may be obtained.

"When simple criteria for evaluating policies do not exist or when performance measures are not evaluated in the same way by all observers, the ability of simulation to provide a total picture of system operating characteristics is a significant advantage over mathematical procedures that produce only single, static answers." (Meier, 1969, p. 22)

Uses

Simulation has proven rather easy to apply both conceptually and practically to engineering problems; but the methodology promises a much higher payoff for management problems, because it enables relationships of a system to be hypothesized and easily tested and because it enables the model builder or manager to exploit his expertise of the system. Simulation has been particularly effective in solving problems relating to inventory control, queuing, scheduling, and resource allocation. However, the uses and application of simulation are varied and diverse as is evident from the following examples.

Jennings and Dickens (1958) modelled the operations of a bus terminal as a combined queuing and scheduling problem. Kuehn and Hamburger (1963) have described a heuristic program for locating warehouses. Richards (1970) has modelled and documented a program for allocation of educational resources in a performanced oriented educational system. Bulkin (1966) developed a simulation for Hughes aircraft for the control of a jobshop operation dealing with local forecasting, labor allocation, and priority sequencing. Steer and Page (1961) developed a computer simulation of the operation of an iron ore unloading port. Smith and Greenlaw (1966) developed a model of the psychological decision process in personnel selection. Kacka and Kirk (1967) developed a large scale computer model which integrated an empirically based model of work groups and foremen with a behavioral theory of the firm. Anderson (1969) has developed a model that deals with the flow of staff allocations in an educational system.

CAMPUS (Comprehensive Analytical Model for Planning in the University Sphere), a model developed by Judy and Levine at the University of Toronto, is one of the best examples of a useful educational simulation model. The model consists of four sections: enrollments, resources, budget, and space. The information flows from one section to another according to specified rules and produces output concerning academic costs, administrative costs, general expenses, and academic and administrative space requirements. It is gaining acceptance as a planning tool for university administrators.

Simulation models have been used to assist in budget preparation, to predict sales fluctuation, to determine student population levels and flow for an educational system, to forecast university needs and operating policies, and most recently to simulate a boxing match between two fighters that could never meet in the ring. New ideas are being found everyday even though some of the users may be making poor use of the technique. As communications improve between technicians and user and as more tools are made available to managers, a great increase in the use of simulation as a problem solving technique should occur. Bibliographies (IBM, 1966; Malcolm, 1960) are available that index simulation literature and can be used to find information concerning specific applications and techniques. Table 2.1 gives some of the current areas in which simulation models are being used. (Emshoff and Sisson, 1970; Kidera and Hoff, 1968)

The basic uses of simulation can be summarized as follows:

1. Simulation for research into the operations of a system.

   This type of simulation is used when the system can be modeled, but solution of the model is either impractical or impossible. It is often possible to simulate the operation repeatedly under various simulated conditions and thus gain insight into the response of the system using various inputs. The Smith and Greenlaw example above could be considered as such a simulation.

2. Simulations to provide data for decision-making.

   This is the most common use of simulation in operation analysis. Many of the studies cited above are of this type. For instance, the Kuehn and Hamburger model for locating warehouses is performed basically so that data can be synthesized in regard to

Table 2.1

Areas In Which Simulation Models Are Currently Being Used

ADDITION OR TRANSFER OF RESOURCES

- Taxi location and dispatching
- Ambulance location and dispatching
- Warehouse location

BEHAVIOR PREDICTION

- Political voting

DESIGN POLICIES

- Urban traffic system design
- Water resources development
- Harbor design
- Information system design
- School attendance area design
- Airport design

FINANCIAL FORECASTING

- Insurance
- Public schools
- Computer leasing
- University operation

LABOR PRODUCTIVITY LEVELS

PERSONNEL POLICIES AND SCHEDULING

- Insurance manpower hiring decisions
- Spacecraft trips

QUEUING POLICIES

- Air traffic control queuing
- Truck dispatching and loading
- Traffic light timing

SALES FLUCTUATION AND CONSUMER BEHAVIOR

- Brand selection
- Advertising allocation
- Promotion decisions

SCHEDULING POLICIES

- Aircraft Maintenance scheduling
- Natural resources scheduling
- Personnel responsibility
- Job shop scheduling
- Public school scheduling
- Bank teller scheduling
- Equipment scheduling
- Railroad traffic scheduling
- Bus (city) scheduling

the best locations. Sometimes a simulation for analyzing the system operation leads into a simulation that will provide data for decision-making. The CAMPUS model is an excellent example of this type.

3. Simulations that are used as training aides.

   This type of simulation is used in management and war games. It is also being used more and more in the training of educators--both administrative and teaching personnel. It is this type of simulation that appears to be making the greatest gains as far as educational use is concerned. It is the opinion of Emshoff and Sisson (1970) that gaming techniques or "in basket" types of simulation will not be successful models for analysis since the use of humans as parts of a simulation introduces too much variability. They do suggest that gaming techniques may find increasing use as artificial realities to help advance research in behavior.

Validation

There is, of course, one sure way to validate a simulation model. Use it for several years and then judge its effectiveness. Since this is not always possible and since users do not like to apply untested models, several approaches have been suggested for "validating" never used models. Meier (1969) recommends a two stage process in validation of models. The first step can be accomplished by making a series of systematic runs of the model to determine whether the model is internally correct in a logical and programming sense. The second step is to determine whether the model represents the real world phenomena it is supposed to represent. This can be accomplished by comparing output from the model with known data from the real world. Meier (1969) further states, "There is no consensus on methods to validate simulation models."

Naylor and Finger (1967) recommend a three stage process in regard to the problem of validation. The first stage requires formulation of a set of postulates or hypotheses upon which the behavior of the model is built. This is a selection process dependent upon the experience and knowledge of the user or model builder since all possible hypotheses cannot be tested. The second stage requires the analyst or user to attempt to "verify" the postulates on which the model is built. The third and final stage consists of testing the model's ability to predict the behavior of the system under study. It is assumed that the final decision concerning the validity of the model should be based on the predictions gleaned from the model.

Conway (1963) outlines another approach to validation. It is concerned with experimentation with the model. He states that this experimentation-validation approach has three phases: model implementation, describing the model in appropriate computer language; strategic planning, designing an experiment that would produce the required information about the system; and tactical planning, deciding how each of the runs that contribute to the experimental design is to be executed. Conway's emphasis is on phase three which has to do with adequate equilibrium and sample size in producing synthetic output from the simulation model.

Hermann (1967) suggest five preliminary tests that can be used in validating never used models. The five tests suggested are as follows:

1. "Internal validity" This is a type of reliability check for the model. Does it have a low variance of outputs when it is replicated with all exogenous (external) inputs held constant?

2. "Face validity" This is simply a test made by asking people concerned with the system under study, whether the model appears to be a reasonable representation of the system.
3. "Variable-parameter validity" This is a type of sensitivity testing. One or more of the factors involved in the model is changed to determine if they affect the output of the model or if they help to make the model produce results that match historical data more closely.
4. "Hypothesis validity" Do the relationships in the model correspond to relationships in the real world? Hermann states, "an operating model would be increasingly valid. . . by evidence of its convergence with the performance of the intended reference system.
5. "Event or time series validity" Does the simulation predict observable events, event patterns, or the variations in output variables?

While there is no general consensus on how to validate a model, there is general agreement that the model should be able to describe the system accurately and should also be able to reasonably predict the future behavior of the system. Positive and absolute "proof" may not be possible concerning a model's validity, but it can be shown that the model has credibility. Emshoff and Sisson (1970) state, "credibility, the only kind of validity we have for a first-time model, requires a detailed examination of the internal structure of the model and of the data used for estimated parameters. It requires careful comparison with such historical data as is available."

CHAPTER III

PROCEDURES FOR CONDUCTING STUDY

This chapter includes a description of the EDSIM 1 simulation model, a description of the system that is to be used in the validation process, and the procedures to be followed in the extraction and coding of data for application of the model.

Description of Model

The EDSIM family of models created at the University of Massachusetts to assist in the planning of individualized performance curriculum consists of EDSIM 1, EDSIM 2, and EDSIM 4.[1] Each of these models traces students through an individualized program, collects data on the simulated students, and produces this data as synthetic output that can be used to answer questions about the overall performance

---

[1]Creation of these models was supported by the Model Elementary Teacher Education Project at the University of Massachusetts, under the direction of Dr. James M. Cooper, primarily as part of the Phase II Feasibility Study performed under U.S.O.E. Contract OEG-0-9-31047-4040 (010) Minor additional support was obtained from General Learning Corporation in conjunction with planning for the Fort Lincoln New Town educational facilities in Washington, D.C.

EDSIM 1 and 2 were created primarily by Mr. George F. Williams and Mr. Wayne E. Leininger, under the guidance of Dr. Eugene E. Kaczka, all of the School of Business Administration, University of Massachusetts.

EDSIM 4 was created by Mr. Thomas C. Richards, as his Ph.D. thesis, under the guidance of Dr. Edward J. Rising of the Industrial Engineering Department, University of Massachusetts.

characteristics of the curriculum being planned and the requirements for and effects of levels of various resources. Each of the models differs in regard to input parameters, number of variables considered, complexity of hypothesized interactions, computer running time, and amount and kind of output data produced.

This study is primarily concerned with the validation of EDSIM 1. The EDSIM 1 model was the first attempt made to trace the progress of individual students through an individualized program. The major variable considered in this model is time.

The model traces a specified number of students through a specified number of randomly selected instructional events in each of several pedagogical areas. An event type is specified by its probability of being selected and the length of time in hours required for its completion. The pedagogical areas, number of performance criteria to be met, and instructional events are specified according to profiles. A profile in EDSIM 1 is defined as the number of performance criteria that need to be mastered in a pedagogical area and the set of instructional alternatives (instructional events) available for each performance criteria. Input data is furnished about each profile in terms of instructional events and the model produces output concerning the following:

1. Student time to completion of program, in event hours.
2. Student time in each pedagogical area.
3. Relative demand for different types of events in different pedagogical areas.

The model requires the following input data for each run.

1. Number of different profiles to be tested (thirteen or fewer).

2. Number of students in each profile.

3. Number of instructional event types available in each profile.

4. For each profile, the number of performance criteria to be met in each of up to thirteen areas of instruction.

5. For each area of instruction:

    a. The probability of passing a pretest and needing no instructional event to satisfy one performance criterion (same for all performance criteria in an area).

    b. The probability of passing a post-test and not needing a second instructional event to satisfy one performance criterion (same for all performance criteria in an area).

6. The time necessary to complete each instructional event type (regardless of the area in which it occurs).

7. For each area, the distribution of event types (probability of selecting each of the available event types to meet a performance criterion in that area for which a pretest was not passed).

8. Identifying comments (forty characters) for output.

After the above input has been specified, each student starts with the given number of performance criteria to be met in each area. For each area the pretest probability distribution is sampled and the number of performance criteria to be met in that area is reduced by the number of pretests successes. It is then assumed that each performance criteria to be met will result in a demand for an instructional event. Next, the post-test probability distribution is sampled the number of times indicated by the number of instructional events to be taken; each post-test failure results in the addition of one more instructional event to be taken.

This number of instructional events which has been adjusted by pretest successes and post-test failures is the number of times the event type distribution for each specific area is sampled to give a number of specific event types for the student. The time spent in each event type is then summed to give a total time requirement for the student in each pedagogical area.

The model then produces the following output based on this total time requirement:

1. The maximum time in event hours for a student to complete work in a given profile.
2. The minimum time in event hours for any student to complete work in a given profile.
3. The mean and standard deviation of time in a profile.
4. A histogram of completion times for a profile.
5. The average time in event hours spent in each area of a profile.
6. A table of the demand for each event type in each area.
7. Percentage of students passing area pretests.
8. Percentage of students failing area post-tests.

In addition to these statistics that are produced internally by the model, it produces the following output based on the input parameters and used for information and filing purposes by the user.

1. Run identification: title and number.
2. Number of students in each profile.
3. Number of students in this run.
4. The probabilities of passing pretests and post-tests in each curriculum area (input by user).
5. The event type probabilities in each curriculum area (input by user).
6. Indication of the specific profile for which this output applies.

Although it could be hypothesized in the real world that completion times, and the probability of event type selection could be a function of student characteristics, no provision is made in this model for such variability. The completion time in the model is considered to be a function of the curriculum and not of student attributes.

Description of System To Be Used in Study

The EDSIM 1 model had been originally developed for use in the development of the Model Elementary Teacher Education Program (METEP) at the University of Massachusetts. Since the METEP program has not been in operation long enough for sufficient longitudinal data to be collected, it could not be used in validating the model. As a result, a system that had similar characteristics and from which time data was available was sought. The Learning Research and Development Center at the University of Pittsburgh made available three years of data from the Individually Prescribed Instruction Program (IPI) in operation at the Oakleaf Elementary School in the Baldwin-Whitehall District in Pittsburgh.

The IPI Program consists of a group of curricula each defined by a series of behavioral objectives, or skills, which are organized into areas and levels. The system requires each student to be placed in each curriculum at a point commensurate with his performance level. Each level involves a sequence of behavioral objectives and each of these defines a criterion of performance. Once a student is placed at the proper point in the sequence, he proceeds at his own rate of progress and demonstrates proficiency in each skill prescribed by his particular instructional sequence. The curricula are constructed so that the objectives within a unit, and from level to level in a given area, are sequenced, because each objective builds on the preceding objective and is prerequisite to those that follow. Individual worksheets designed to develop particular skills are the primary instructional activity used in the system and are available to students at all times.

In order to monitor pupil progress and furnish necessary information to curriculum developers, test constructors, and teachers, IPI employs a very comprehensive testing program. The testing program includes Placement Tests, Unit Pretests, Unit Post-tests, and Curriculum-Embedded Tests. A proficiency level of eighty to eighty-five percent has been established for all tests used in the IPI system.

Placement tests are administered at the beginning of a school year or, for a new pupil, when he enters school. This test is designed to give the teacher information about the student in regard to the unit he might be assigned. Specific information about the objectives or skills within a unit are furnished by the Unit Pretest. This test measures performance on each objective in the unit so that the teacher has necessary information for specific assignments in the Unit. A student may pretest out of specific objectives for a unit or may pretest out of an entire unit of work as a result of content learned at lower level or in a related unit. The pretest is administered whenever a student is about to enter a unit.

As a pupil proceeds through specific objectives in the units, there is need for evaluation of his performance. Two types of tests are used for this purpose. The first of these is the Curriculum-Embedded Test (CET). This test provides information concerning pupil progress from skill to skill. A CET is prescribed for the pupil as the final exercise in his set of instructional materials for each objective. On certain CET's there has been added a second part which measures pupil performance on the next objective in the unit. This provides a type of short pretest for those pupils whose instructional sequence includes a prescription for that objective. If this CET pretest shows that the student has been able to transfer or generalize his learning to gain command of the next skill, work on that skill may be omitted.

The second type of test used to evaluate student performance is the Unit Post-test. This test is used for reassessment of performance on the unit as a whole.

Evaluation of the pupil's performance at this point is required to determine whether he moves to the next unit or continues work in the present unit. Success on the post-test indicates that he is ready for the pretest for the next unit; failure indicates that he needs more instruction in the present unit. All tests in the IPI program are conducted as aids to assist teachers in making decisions concerning prescriptions for students; however, the program allows the teacher discretionary power to add additional cycles or to pass the student on to the next unit even if the score on a test indicates the contrary.

Since it is necessary to monitor individual students for comprehensive teacher information, IPI maintains a comprehensive information system in machine readable form. Data from this information system comprised the data tapes made available for the validation study. It was evident from examination of the available data tapes that data from the IPI system included components analogous to those proposed in METEP.

1. Pretesting -- to determine what, if any, instruction is needed to master a given area or skill.

2. Designation of the instruction needed and recording of the time needed to complete an instructional prescription resulting from a pretest.

3. Post-testing -- to determine proficiency after the instruction.

4. Recording of additional time needed if proficiency has not been reached.

5. Total time to completion of a segment (unit) of a particular level.

There are, of course, differences that exist between the IPI system and METEP. The major differences are the following:

1. The METEP student may select a particular subset of skills in which to demonstrate or acquire proficiency.

   The student in IPI has a relatively linear set of skills to master. The individual differences of the IPI student are evident only in the rate at which he progresses or in the amount he masters before he leaves the system.

2. The METEP student selects the total amount of time he wishes to spend in the system.

   The IPI student receives an hour of instruction in each curriculum area (reading, mathematics, _etc_.) each day.

3. The METEP student has a wider range of activities available to him and may be involved in activity scheduling or waiting time which could affect his rate of progression in the program.

   The IPI student has almost every instructional alternative available to him at all times since pre-printed worksheets are the primary instructional activity.

4. The amount and kind of staff that each of these students might require could be very different.

However, all differences considered, the IPI data offer one of the best opportunities available for a validation study of the EDSIM 1 model, especially since longitudinal data is available in machine readable form. The data from the Oakleaf School was received on magnetic tapes.

The data from the Mathematics Curriculum is used in this study. The present IPI Mathematics Curriculum is divided into content area of numeration, place value, addition, subtraction, multiplication, division, combination of processes, fractions, money, time, systems of measurement, geometry, and special topics, and extends over levels A through H. Table 3.1 displays the content areas, levels, and number of objectives involved in each.

Examination of the data revealed three possible levels of detail that could be used in describing IPI math data for EDSIM 1 model validation.

1. Make each level in the math curriculum a "unit" and simulate completion of the entire math curriculum. This approach was rejected since close examination of the data revealed that no one had completed the entire curriculum. This approach would also have proved difficult since practically no one could pretest out of an entire level. It is also difficult to determine an appropriate post-test for an entire level, and this approach would have made it necessary to exclude this feature from the model.

2. Make each "unit" the skills within an area of a specific level. This approach would have meant considering CET's as post-tests and that type of detail is difficult to extract from the data available. The time variability in the skills level is very small and depends more on the number of skills prescribed rather than on the time to complete one skill. A further problem would have arisen in attempting to run the model at this level of detail. Occasionally, a student will have prescribed worksheets from a previous level as review or reinforcement material while working in a higher level; i.e., Level D numeration material while working in Level E numeration. It is possible to account for this time as service time in Level E numeration but difficult to account for the time as specific skills in Level E numeration.

3. Make each "unit" the intersection of an area and level. Pretests, post-tests, and service time distributions can be more accurately described at this level.

The third approach concerning the level of detail to be used in this study has been chosen as a good compromise between sufficient

Table 3-1

Number of Objectives for Each Unit in the IPI Mathematics Curriculum

CONTENT AREAS	A	B	C	D	E	F	G	H
Numeration	12	10	8	8	8	3	3	4
Place Value		3	5	10	7	5	2	1
Addition	3	10	5	8	6	2	3	2
Subtraction			4	6	3	1	3	1
Multiplication				8	11	10	6	3
Division				7	7	9	5	5
Combination of Processes			6	5	7	4	5	6
Fractions	3	2	4	6	6	14	5	2
Money		4	4	6	4	1		
Time		3	2	7	9	5	3	1
Systems of Measurement		4	3	5	7	3	2	
Geometry		2	2	3	9	10	7	9
Special Topics			1	3	3	5	4	5

detail to the features of the time component of EDSIM 1 and the practicalities of managing and interpreting the data available. For the purpose of this study the following definition of terms is used.

Level: A major division of the IPI Mathematics Program. Eight levels of work-difficulty comprise the math program (levels A through H). This study will focus on levels B, C, D, and E.

Area: One of the thirteen subject matter topics offered in a given level (numeration, place value, addition, *etc*.).

Unit: The group of skills that comprise the area of a given level. A unit is specific for one area in a level. In Table 3.1 , a unit is the cell at the intersection of an area and level. In Table 3.1, the skills (8) that must be completed under Level E--Numeration will be treated as a performance objective in applying the model. For ease of reporting in tables, Level E--Numeration will become E1, Level E--Place Value will become E2, *etc*. The same type of coding will be used for Levels B, C, and D.

In applying the IPI data to the EDSIM 1 model, the following relationships shall exist:

IPI TERM	METEP TERM
Level	Profile
Area	Pedagogical Area
Unit	Performance Objective
Prescription	Instructional Alternative

Data Preparation

The data from Oakleaf was received on magnetic tape. Preliminary examination of the data revealed that certain editing would be necessary before it could be used for the purpose of this study. The 1967-1968 and 1968-1969 tape includes data from the Mathematics and Reading components as well as pertinent background data. The 1969-1970 tape includes background and Mathematics data. The background data includes scores of IQ tests, placement tests, etc. and was placed in its own file for use in determining relationships between student attributes and time spent in the system. Further discussion of this data appears in a later section of this study.

In order to prepare an operational file of Mathematics component data to be used in the validation study, the following editing procedures were performed.

1. Mathematics data was identified and placed on tape for further editing.
2. This file of data was examined for missing information and obviously incorrect data. Each record was searched to determine if it included the following information: level, area, student number, data of recording, and record type (pretest, post-test, or precription data). Records that did not include all aspects of this information were discarded. This resulted in elimination of between one and two percent of the data.
3. This tape file was then searched for duplicate records. All duplicate records were removed from the file. The resulting tape was then repositioned into a fixed format so that it could be sorted by transaction date (recording date) and record type to obtain chronological and time distribution data for each student.

The resultant tape file is the operational file used to extract time distribution data for use in the EDSIM 1 model. The file is segmented by school year and contains all retrievable and identifiable information on students doing any work on Levels A through H of the IPI

Mathematics component. Within each type of record the appropriate test score or time spent on a skill is recorded.

In order to obtain the parameters for EDSIM 1, several assumptions have been made in processing this operational tape file. These provide de facto definitions of the parameters used in this validation study. Detailed explanations of these assumptions follow.

Student sequence of units: For each student a sequence of units was defined. This begins each year with the first unit in which the student receives instruction and ends with the last unit for which activity is recorded. A student is presumed to pass through all the units included in the sequence, either pretesting out of a sequence for a student could be D8 through E9. This means that the student either participated in or pretested out of units D8, D9, D10, D11, D12, D13, E1, E2, E3, E4, E5, E6, E7, E8, and E9, a total of 15 units.

Pretest: This may be a formal diagnostic test at the beginning of the year, a specific unit pretest, or the simple fact that a student apparently did not take instruction in the next unit in his sequence.

Probability of passing a pretest: The number of students who participated in a specific unit was summed and divided by the number of students that had that specific unit in their sequence. This ratio was subtracted from 1.0 to determine the probability of passing a pretest.

Post-test: A student may pass a formal post-test for a unit or may move on to another unit at his teacher's option if he is having difficulty over a long period of time. A post-test may be taken more than once.

Probability of passing a post-test: The first identified post-test is taken as the post-test to be considered. In some cases this is nothing more than progression to work in the next unit. In most cases, however, it is a formal, identified test. The number of students that take more work in a unit following a post-test is summed and divided by the number of students taking any instruction in a specific unit. This ratio is subtracted from 1.0 to determine the probability of passing a post-test.

Service Time: The time in days (or hours, since math is allotted one hour per day) for each student in each unit prior to the first identified post-test and subsequent to that post-test is calculated. The pretest to post-test service time includes one day for pretesting and one day for post-testing. For each student in each unit a distribution of service time is produced for use in the EDSIM 1 model.

This information concerning time before and after a post-test was produced by a computer program for each student that participated in the Mathematics component each year and was printed out for easy reference and examination. Table 3.2 is an example of the data produced for one student in the 1967-1968 school year. This information was also punched into cards in a fixed format for easy handling in making frequency counts and service time probability distributions for use in the model. Table 3.3 is an example of the time distribution data for a sample of thirty-eight students that completed Level E in 1967-1968.

The service time distributions were calculated by forming frequency distributions for each unit in a level. The frequency distributions were set up by event types (ET) (defined on page) and number of students that selected (required time in) that event type. In Table 3.3 it can be observed that thirty-four students spent time in Unit 1 before the post-test and twenty-seven spent time in the unit after the post-test. Of this total of sixty-one, six students required seven days (ET 7) of work in the unit (four before the post-test and two after the post-test). The probability that a student takes seven days in this unit is 6/61 or .098. A simple FORITAN program was written to take the information punched in the fixed format data cards and produce the above information for each ET.

It was necessary that each of these probability distributions be checked since rounding off procedures produced distributions with a

Table 3-2

Example of Time Distribution Data Produced for Student 0055 in 1967-1968

TIME SPENT BY STUDENT NO: 0055

1967-1968

Before Post Test No.: 1			After Post Test No.: 1		
LEVEL	AREA	TIME IN DAYS	LEVEL	AREA	TIME IN DAYS
D	1	3	D	1	4
D	10	3	D	10	2
D	11	5	D	11	3
D	12	1	D	12	0
E	1	7	E	1	0
E	5	11	E	5	0
E	6	3	E	6	2
E	7	4	E	7	3
E	8	9	E	8	2
E	9	1	E	9	0
E	10	8	E	10	8
E	11	14	E	11	6
F	4	1	F	4	0
F	6	26	F	6	3

Table 3.3: Time Distribution Data for Sample of 38 Students That Completed Level E in One School Year.

	TIME BEFORE POST TEST													TIME AFTER A POST TEST															
	UNITS													UNITS													TOTALS		
	1	2	3	4	5	6	7	8	9	10	11	12	13	1	2	3	4	5	6	7	8	9	10	11	12	13	B	A	T
0011 E	17	7	0	10	20	21	5	7	2	0	0	0	0	6	0	0	3	1	4	10	0	1	0	0	0	0	89	25	114
0033 E	7	0	0	0	0	0	0	0	2	0	9	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	18	1	19
0044 E	7	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	10
0055 E	7	0	0	0	11	3	4	9	1	8	14	0	0	3	0	0	0	0	2	3	2	0	8	6	0	0	57	24	81
0066 E	18	0	4	11	25	12	13	9	1	8	19	6	0	11	0	9	10	4	2	0	5	0	0	3	3	0	126	47	173
0077 E	13	0	0	4	16	13	0	5	5	0	4	0	0	17	0	0	1	12	2	0	0	5	0	7	0	0	60	44	104
0113 E	6	0	0	0	1	1	3	0	1	0	4	1	0	5	0	0	0	0	0	1	0	0	0	1	0	0	17	7	24
0124 E	10	0	0	0	18	9	11	17	4	16	9	2	0	1	0	0	0	7	1	9	1	6	8	0	10	0	96	43	139
0135 E	11	6	0	0	12	0	5	2	1	15	2	0	0	11	0	0	0	2	0	0	2	0	0	4	0	0	54	19	73
0146 E	11	0	0	0	0	0	0	0	1	7	18	10	0	10	0	0	0	0	0	0	0	0	4	11	0	0	47	25	72
0179 E	13	0	0	0	16	1	8	19	1	17	12	2	0	2	0	0	0	0	0	0	8	0	0	11	2	0	89	23	112
0181 E	10	8	4	0	18	10	2	6	1	8	14	2	0	15	4	1	0	16	0	15	4	0	7	11	5	0	83	78	161
0192 E	6	0	0	6	0	12	0	7	5	12	0	0	0	13	0	0	6	0	0	0	0	4	2	0	0	0	48	25	73
0215 E	9	6	0	7	0	0	9	9	4	0	9	12	0	10	0	0	0	0	0	2	4	6	0	17	7	0	65	46	111
0237 E	10	0	0	0	4	9	0	0	0	14	13	9	0	0	0	0	0	6	11	0	0	0	0	10	1	0	59	28	87
0248 E	14	0	4	0	16	11	8	2	6	18	8	9	0	5	0	0	0	0	8	15	7	6	1	0	4	0	96	46	142
0259 E	17	0	0	0	11	19	13	2	2	0	11	13	0	0	0	0	0	20	0	5	9	5	0	0	3	0	88	42	130
0272 E	12	13	0	0	1	25	10	0	1	5	29	0	0	1	0	0	0	0	0	2	0	0	12	3	0	0	96	18	114
0294 F	11	5	0	0	13	0	0	12	2	0	8	3	0	1	1	0	0	23	0	0	3	1	0	7	7	0	54	43	97
0328 E	8	0	0	0	11	10	2	4	6	5	12	10	0	6	0	0	0	1	0	2	0	0	0	0	9	0	68	18	86
0339 E	6	0	0	8	2	16	5	12	7	0	0	0	0	0	0	0	6	8	1	3	7	0	0	0	0	0	56	25	81
0341 E	10	0	0	4	13	0	5	12	5	12	2	0	0	5	0	0	0	0	0	3	3	0	8	5	0	0	63	24	87
0363 E	0	0	12	12	24	0	0	11	0	5	2	6	0	0	0	0	5	0	0	0	0	0	4	3	8	0	72	20	92
0421 E	10	0	0	0	1	1	1	1	1	0	7	9	0	2	0	0	0	0	0	0	0	0	0	0	3	0	31	5	36
0432 E	6	3	0	0	8	7	8	0	0	0	0	0	0	7	2	0	0	0	0	0	0	0	0	0	0	0	32	9	41
0454 F	4	0	0	1	11	0	0	0	3	0	10	0	0	7	0	0	0	1	0	0	0	0	0	0	0	0	29	8	37
0487 E	28	6	0	17	0	0	0	0	0	0	0	15	0	12	0	0	1	0	0	0	0	0	0	0	0	0	66	13	79
0498 E	2	0	12	0	0	0	0	0	0	22	0	17	0	14	0	0	0	0	0	0	0	0	0	0	3	0	53	17	70
0512 E	12	0	0	13	21	15	0	7	0	10	15	12	0	0	0	0	8	0	0	0	4	0	3	2	2	0	105	19	124
0556 E	8	0	0	18	27	12	0	6	0	0	17	2	0	3	0	0	9	10	0	0	4	0	0	9	9	0	90	44	134
0567 E	0	0	0	5	18	5	0	1	0	0	2	2	0	0	0	0	15	0	0	0	0	0	0	19	9	0	33	43	76
0578 E	2	0	0	9	0	19	0	0	0	0	0	13	0	5	0	0	0	0	1	0	0	0	0	0	3	0	43	9	52
0614 E	7	0	6	3	0	4	9	3	4	11	12	9	0	1	0	0	11	0	10	0	3	2	1	0	5	0	68	33	101
0738 F	10	0	0	0	3	22	16	10	6	11	3	0	0	10	0	0	0	0	0	0	3	1	13	1	0	0	81	28	109
0818 E	0	0	0	6	26	32	10	6	7	0	14	3	0	0	0	0	4	6	0	0	5	1	0	5	4	0	104	25	129
0875 F	6	5	0	0	13	18	0	10	4	9	1	9	0	1	0	0	0	0	0	0	0	1	2	0	2	0	75	6	81
0897 E	0	0	0	0	0	0	18	15	0	3	13	11	0	0	0	0	0	0	0	2	2	0	5	7	0	0	60	16	76
0922 E	9	0	0	0	2	18	2	15	1	6	2	2	0	0	0	0	0	4	2	5	3	0	0	6	4	0	57	24	81

cumulative probability above or below 1.00. Any total above or below 1.00 leads to malfunction of the model. If the probability adds only to .98 for a unit distribution and the true range of selection is from ET 1 to ET 15, ET 25 (the last Event Type) has a probability of .02 of being selected. This type of input error is easily found in the model. By checking the Demand for Event Types Table produced by the model, it can be observed if ET 25 (or last Event Type) has been selected when its probability of selection was input as zero. If it has, input distributions should be reexamined. Incorrect input of this type can greatly affect the overall output of this model since it leads to selection of extreme times inconsistent with system operation and with the true range of time in a unit.

The same type of error can be produced if the cumulative total is above 1.00. If the cumulative probability is 1.02 for a unit distribution and ET 18 is the last Event Type in the unit time range with a selection probability of .03, its probability of selection is reduced to .01. This is a more difficult type of error to find. However, if in iteration of the model with different series of random numbers, it is observed that ET 18 is never selected as many times as other Event Types with comparable probabilities, input should be reexamined. This type of error does not affect output very substantially if the range of time in a unit is narrow; i.e., from three to ten days. However, it can have a great affect on output for units that extend over a wide range since it reduces the occurances of events of longer duration and reduces the probability of selecting an extreme event.

CHAPTER IV

DATA ANALYSIS

Included in this chapter are the reports of the various tests to which the model was subjected. Tests were run to determine the affect of the random number generator on the model. Sensitivity tests were conducted to determine the model's ability to react to small differences in input parameters. Descriptive tests were run to determine the model's ability to describe time to completion of a level both within a year and across years. After improvements were made to the model, statistical tests were made to determine the improvement in the descriptive ability of the model. The final tests made on the model were tests to assess its ability to predict the time to completion of a level.

The Kolmogorov-Smirnoff Two Sample Test was used to determine the statistical difference in the simulated and observed distributions. Graphs of the simulated and observed distributions were drawn so that comparison of these distributions could be visually displayed.

In order to use the model with the IPI input parameters, changes were made in the labeling features of the output section of the model so that the output produced would be labeled in relation to the IPI system. This is an important aspect in the validation study since great amounts (unbelievable, in fact) of output data can be accumulated very quickly in the use of any simulation model and explicitly labeled output is a necessity.

The original model had labeled output in relation to Profiles, Pedagogical Areas (Language Arts, Mathematics, Science, *etc*.), and Performance Objectives. The format statements in the original model were changed so that the output would be labeled in relation to Levels, Areas (Numeration, Place Value, Addition, *etc*.), and Units. This required no change in the logical operation of the model.

The model with changed output formats was then tested using input parameters from a sample of thirty-eight students that had completed Level E in 1967-1968. The output from this preliminary run was examined carefully to see if obvious coding errors were evident. Examination revealed that the last event type in every Unit was being selected even though its probability of selection according to input data was zero. Examination of the programming in the model revealed that the cumulative distribution being built in the model for sampling of service time distributions by use of random numbers had been coded to respond to a particular type of random number generator. Since the random numbers to be used in this study were to come from the Random Number Generator (Ranf-1) of the CDC 3600 Computer at the University of Massachusetts, it was necessary to make a correction in this aspect of the model. This correction was made.

Test runs were made again and all aspects of the model's output were examined. Manual calculations were done on certain arithmetical aspects of the model to see if other obvious errors could be found. No other errors could be found. This operational model with changed output formats is the operational source model on which this study is based. A sample of the output from this model is presented in Figure 4.1.

PAGE NO. 1

E D S I M I

S C H O O L O F E D U C A T I O N S I M U L A T I O N

RUN NUMBER 2

DATA DERIVED FROM EDSIM I LEVEL E IPI 67-68 ESAMP 11-70

THE NUMBER OF STUDENTS IN EACH LEVEL IS 144

THE NUMBER OF LEVELS FOR WHICH RESULTS WILL BE OUTPUTED IS 1

PRE AND POST TEST PROBABILITIES FOR THIS RUN

	NUM.	P. V.	ADD.	SUB.	MULT.	DIV.	C.O.P.	FRACT.	MONEY	TIME	MEAS.	GEO.	SP.TOP.
PRE	0.080	0.770	0.840	0.580	0.260	0.320	0.420	0.260	0.260	0.420	0.210	0.340	*****
POST	0.400	0.670	0.670	0.250	0.460	0.580	0.360	0.320	0.540	0.360	0.300	0.160	*****

Figure 4.1: Sample Output from the EDSIM I Model After Labeling Formats Were Changed

Random Number Generator Check

Before the model could be tested as to its ability to describe the system under study, tests were made to determine the effect of the random number generator on the model. A number of tests have been developed to determine whether a series of random numbers meets the criterion of randomness. A generator that produces numbers for use in a simulator such as EDSIM I is producing random numbers if any of the range of numbers is equally likely to occur and if each new number is completely independent of any previous output of the generator. Statistically, this means that the numbers are uniformly and randomly distributed.

Emshoff and Sisson (1970) recommend that random number generators be tested only on characteristics of randomness most critical to the accuracy of the implications that will be drawn from the model rather than on every randomness criterion for which tests are available.

The EDSIM I model uses random numbers to determine whether a student passes a pretest or a post-test. If the probability of passing a pretest is .16, any number produced by the generator between zero and .15 determines a pass and any number larger than .16 determines a fail. The same procedure is used in the post-test sampling.

The model also uses random numbers in the selection of event types. Service time distributions are determined in the model by fitting a random number into a cumulative distribution of service time event types. If Event Type (ET) 1 had a probability of selection of .15; ET 2, .45; ET 3, .10; ET 4, .05; ET 5, .15; and ET 6, .10; a cumulative distribution of event types is built within the model by adding each event type probability to the next to produce areas of selection. This would make the cumulative distribution .15, .60, .70, .75, .90, 1.00. If the random number generator then produced the number .34, Event Type 2 would be selected.

In order to test the ability of the generator to produce numbers that would allow equal opportunity for each of the twenty-five event types to be selected, the model was run with uniform probabilities (.04) in each of the twenty-five event types. Two hundred random numbers were generated in each run of the model and ten runs of the model were made. The pretest probability was set at zero and the post-test probability at 1.00. This caused the unit distribution in each area (numeration, addition, etc.) to be sampled 200 times. The number of times an event type was selected in each of the distributions was recorded and punched into a fixed format on cards. If these service times distribution were truly uniform (every event type has equal ability to be selected) each of the twenty-five event types should have been selected eight times. These runs of the model had produced 130 replications of unit service time distributions (thirteen Areas in each of the ten runs). Table 4.1 shows a sample of these service time distributions from run number 2. In order to test how significantly these distributions differed from a truly uniform distributions, a Chi Square Goodness of Fit Test was administered to each of these distributions. A simple FORTRAN program was written to take the information from the punched cards and conduct the test.

Table 4.1: Sample of Uniform Distribution Check for Unit Service Time Distributions from Run Number 2 Using the CDC 3600 Random Number Generator (RANF-1).

PAGE NO. 4

DEMAND FOR EVENT TYPES-WITH 25 TYPES AVAILABLE

	NUM.	P. V.	ADD.	SUB.	MULT.	DIV.	C.C.P.	FRACT.	MONEY	TIME	MEAS.	GEO.	SP.TOP.
ET 1	11	12	13	7	5	8	9	7	12	8	7	9	13
ET 2	8	7	10	7	8	6	5	6	4	8	10	6	9
ET 3	5	11	5	9	8	10	7	7	5	7	11	8	4
ET 4	7	14	7	3	4	7	9	8	8	5	11	11	6
ET 5	8	8	11	6	7	8	4	6	12	6	5	10	7
ET 6	9	8	8	5	9	9	10	7	7	11	6	1	7
ET 7	7	6	13	10	9	15	6	6	6	8	8	8	8
ET 8	6	2	9	14	11	8	10	10	8	9	5	8	5
ET 9	10	7	10	5	9	8	8	8	12	12	10	2	9
ET10	5	7	3	9	9	7	8	9	7	12	7	13	7
ET11	11	10	5	9	10	6	5	6	6	7	11	6	5
ET12	5	9	10	2	18	10	9	9	7	7	13	4	8
ET13	8	5	8	9	6	10	4	7	8	8	2	11	8
ET14	9	10	4	4	4	10	9	8	7	4	6	15	14
ET15	8	9	4	7	9	5	8	6	6	6	13	7	7
ET16	10	4	4	11	9	8	8	6	6	5	4	11	10
ET17	9	10	6	7	9	5	7	5	10	12	12	7	9
ET18	5	14	7	9	5	11	7	10	10	10	9	7	7
ET19	7	6	8	12	4	5	13	12	7	5	8	8	12
ET20	6	12	12	7	7	9	4	13	8	10	8	5	7
ET21	10	6	7	5	6	9	9	13	5	4	11	8	11
ET22	8	6	7	12	12	6	6	7	10	8	4	5	7
ET23	9	4	9	11	8	9	12	3	9	5	6	10	6
ET24	10	5	13	13	8	5	11	13	9	10	6	12	6
ET25	9	8	7	7	6	6	12	8	11	13	7	8	8

SUM OF ABOVE DEMAND - BY TYPE

121	94	97	100	98	97	110	105	110	103
97	111	94	104	95	96	108	111	107	108
104	98	101	121	110					

In conducting the Chi Square Test the following hypotheses were tested.

H_0: there is no difference in the number of times an event type is chosen and any observed differences are chance variations to be expected in a random sample from a uniform distribution where $f_1 = f_2 = f_3 = \ldots\ldots\ldots\ldots f_{25}$.

H_1: the frequencies $f_1, f_2, f_3, \ldots\ldots f_{25}$ are not all equal.

These hypotheses were tested against a significance level of .05 with degrees of freedom = k - 1 = 25 - 1 = 24.

In order to reject H_0 at the .05 level the value of X^2 would have to be greater than 36.42. In the 130 relications at this significance level one might expect six or seven distributions to be significantly different. Only one of the distributions had a X^2 greater than 36.42. The table of X^2's produced is in the appendix.

Since it would be possible to have non-significant X^2's and yet have all of the first thirteen event types be below the expected and all of the last twelve event types be above the expected a simple runs test was made on the 130 service time distributions produced in the random number check. This technique is based on the number of runs that a sample exhibits. A <u>run</u> is defined as a succession of identical symbols which are followed and preceded by different symbols or no symbols at all. The data in Table 4.2 is a sample of this type of test. It can be noted from Table 4.2 that a plus (+) symbol had been placed above a frequency that is above the expected (8) and a minus (-) symbol is placed above those frequencies that are below the expected. The number of runs and the number of symbols of each type is then counted.

Table 4.2

A Sample of the Runs Tests Made on the 130 Replications of Unit Service Time Distributions

	FREQUENCIES IN EACH OF THE TWENTY-FIVE EVENT TYPES																								
	+	-	-	+	+	+	+	-	-	-	+	-	+	-	+	-	-	-	+	-	-	-	+	+	-
NUMERATION	9	6	7	10	13	9	9	7	3	7	9	3	13	7	13	5	7	6	9	6	7	7	1	10	7
	1	2		3				4			5	6	7	8	9	10			11	12			13		14

r = runs = 14

N_1 = plus = 11

N_2 = minus = 14

r must be less than 8 or greater than 18 to be significant at the .05 level

The hypotheses to be tested are:

H_0: the pluses and minuses occur in random order

H_1: the order of the pluses and minuses deviates from randomness

In the example presented in Table 4.2, it can be observed that $N_1 = 12$, $N_2 = 13$, and r = runs = 13. This information is then checked against a Table of Critical Values of r in the Runs Test (Seigel, 1956) to see if the r is significant. The table indicates that an r equal to or less than eight or equal to or greater than twenty is required to reject the null hypothesis at the .05 level (the table only presents information for the .05 significance level). None of the 130 replications of the service time distributions checked in the same way against the table were significant.

The results of these two tests lead to the conclusion that the random number generator being used produced numbers that could meet a randomness criterion for use in EDSIM 1. There appeared to be very little bias that could be contributed to the random numbers being used.

Sensitivity Tests

The next tests run on the model were tests to determine the ability of the model to react to small differences in input parameters. Initially two groups of students were selected. The first group were all thirty-eight students that completed Level E Mathematics in the 1967-1968 school year. The second group were all 103 students who participated in any part of Level E Mathematics during that school year (including the first sample of thirty-eight students).

Table 4.3 shows the pretest and post-test probabilities, and the distribution of service times for both samples of nine of the units, under the assumption that the service times before and after the first identified post-test may be combined into one distribution. Event Types 1 through 21 each represent that number of days. Event Types 22 through 25 represent the mid-point of an interval of days. Event Type 22 takes twenty-three days, the mid-point of twenty-two to twenty-four days, and similarly, Event Types 23, 24, and 25 represent twenty-six, twenty-nine, and forty days respectively. The number of days, hours, etc. that each of these event types represents is set by the user after preliminary examination of the data. The event types are input for each Level. The above event types were found to be realistic for all Levels except Level B. In simulating Level B, it was necessary to change Event Types 24 and 25 to thirty-five and fifty days, respectively so that extreme times spent in some of the units in this Level could be better represented. These data in Table 4.3 are used to test the model's ability to replicate the distribution of times required to complete Level E Mathematics from the input data.

It can be observed that there are small differences between the two samples in pretest and post-test probabilities as well as in the service time distributions. In order to determine how much these differences would affect the output of the EDSIM 1 model, a computer run was made using each of these sets of data as input. Each computer run consisted of ten samples of thirty-eight students each; each computer run used the same sequence of random numbers generated by the University

Table 4.3

Sample of Input Data for Nine Units for Two Samples of Students in Level E Showing Probability of Passing Pre Test, Probability of Passing Post Test, and Service Time Distribution by Unit

	UNITS 1		2		3		4		5		6		7		8		9	
	P	C	P	C	P	C	P	C	P	C	P	C	P	C	P	C	P	C
PRE TEST PROB	.16	.08	.52	.77	.70	.84	.46	.58	.26	.26	.30	.32	.43	.42	.32	.26	.55	.26
POST TEST PROB	.35	.40	.55	.67	.40	.67	.32	.25	.33	.46	.54	.58	.34	.36	.30	.32	.40	.54
EVENT TIME 1	.01		.04		.04		.11	.07	.05	.11	.06	.11	.05	.05	.09	.11	.31	.39
EVENT TIME 2	.12	.06	.09		.04		.06	.06	.17	.07	.11		.05	.13	.09	.11	.14	.14
EVENT TIME 3	.02		.04	.11	.08	.50	.02		.05	.04	.04	.04	.05	.05	.03	.04	.06	.04
EVENT TIME 4	.01	.03	.04		.16		.08	.13	.02	.03	.06	.04	.03	.05	.04	.03	.17	.14
EVENT TIME 5			.11	.22	.12		.08	.06	.02		.04	.04	.13	.17	.06	.03	.11	.11
EVENT TIME 6	.07	.15	.18	.33	.12	.17	.12	.13	.02				.03		.06	.11	.08	.10
EVENT TIME 7	.09	.11	.07	.12	.04		.08	.06			.06	.04	.11		.09	.11	.11	.07
EVENT TIME 8	.02	.06	.09	.11	.04		.06	.06	.03	.03	.02		.14	.13			.02	
EVENT TIME 9	.05	.06			.04		.08	.06			.04	.08	.11	.09	.09	.11		
EVENT TIME 10	.10	.17	.04		.04		.11	.06	.02		.05	.07	.11	.09	.06	.07		
EVENT TIME 11	.07	.09	.03		.12		.06	.06	.07	.14	.02	.04	.03	.05	.05	.03		
EVENT TIME 12	.02	.06	.04		.08	.33	.04	.06	.03	.03	.05	.11			.14	.11		
EVENT TIME 13	.10	.06	.07	.11			.02	.06	.10	.11	.05	.04	.08	.09				
EVENT TIME 14	.07	.03	.07				.02		.08						.09			
EVENT TIME 15	.04				.04					.11	.05	.04	.03			.07		
EVENT TIME 16	.01		.03		.04						.04	.04		.05	.05			
EVENT TIME 17	.07						.02	.06		.11	.02		.05		.03	.03		
EVENT TIME 18	.02	.06					.02	.07			.06	.07		.05				
EVENT TIME 19	.01	.03	.03						.06	.04	.04	.08			.03	.04		
EVENT TIME 20	.01		.03						.02	.04		.04						
EVENT TIME 21	.01								.05		.04	.04						

Table 4.3 (Continued)

	UNITS 1		2		3		4		5		6		7		8		9	
	P	C	P	C	P	C	P	C	P	C	P	C	P	C	P	C	P	C
EVENT TIME 22	.05								.03	.07	.09	.04						
EVENT TIME 23	.01								.10	.07								
EVENT TIME 24	.01	.03							.03		.06							
EVENT TIME 25	.01								.05			.04						

P = Sample of 103 students that participated in but did not necessarily complete Level E in one year

C = Sample of thirty-eight students that completed Level E in one year of Massachusetts CDC 3600 System Random Number Generator to keep the effects of the random number generator constant.

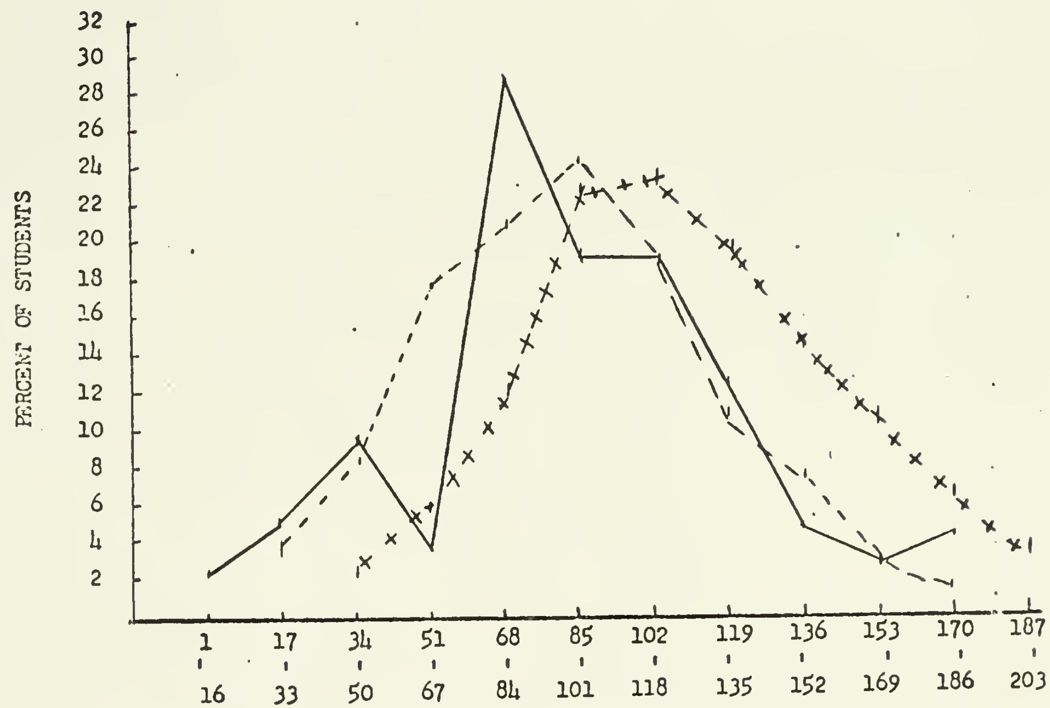
The output from each computer run was combined into a frequency distribution of the percent of students falling into each of several time intervals for completion of Level E Mathematics and compared to the actual completion time distribution for the first group of thirty-eight students. Figure 4.2 displays the results.

Several comments may be made about the distributions displayed in Figure 4.2. First, thirty-eight is a fairly small sample size, and since it represents the totality of available real data, it cannot readily be "smoothed" by combining with other samples. Consequently, somewhat more deviation of observed from predicted distributions may be expected. Second, it appears that using data from the 103 students (although not all 103 affected every input parameter) gives a distribution indicating somewhat longer completion times than was obtained from using parameters based on the thirty-eight students who actually completed the entire Level E. The distributions do indicate, however, that the EDSIM 1 model is sensitive to relatively small differences in input parameters.

Descriptive Tests

The next tests of the model were tests that would assess the model's ability to describe the system in terms of the variables and parameters used in the model. Samples from Levels B, C, D, and E were taken from the 1967-1968 data. These samples were used for estimating

FIGURE 4.2: Comparison of Observed Distribution to Simulated Distribution Using Input from 38 Students That Completed Level E and Input from 103 Students That Participated in Level E.



———— Observed Data ------- Simulated Distribution (Input from 38 Students)

+++++++++ Simulated Distribution (Input from 103 Students) parameters for the variables in the model and included all the students that had participated in each of these levels. If a student participated in levels B, C,and D during that year, his time distribution data was sorted by level and included in the sample for that level. There were seventy-six in the Level B sample; 116 in the Level C sample; 144 in the Level D sample; and 103 in the Level E sample. Since one student's data could be included in more than one sample, the total of these samples (439) does not represent the total students that were in the IPI system that year.

The second sample of students selected were samples of students that had completed a specific level in that year. There were thirty-six students in the sample from Level B; fifty-six in the Level C sample; seventy in the Level D sample; and thirty-eight in the Level E sample. These samples were used to construct the frequency distribution with which the simulated samples would be matched. Parameters for service time distributions and pretest and post-test probabilities were also derived from these samples. These parameters were compared with the parameters taken from the "participating" samples.

It was observed from these comparisons (an example for Level E is presented in Table 4.7) that the probabilities for passing the pretest were usually lower for those from the sample that completed the level but the probability of passing the post-test were higher for the "completing" sample. This type of trend lead to the assumption that some of the students in the "completing" sample were not encountering the material for the first time but could be using it as review or reinforcement material.

This would mean that they had not been placed in a specific unit because of a pretest but had been placed there at the discretion of the prescriber. This is not to say that any conclusion could be drawn but only that the sample that was to be used as the observed distribution of service times might not be a true representation of the sequence that was an integral part of the curriculum construction.

It was also observed that service time distributions for the "completing" sample consistently had a smaller range of days than the "participating" sample and also higher probabilities in the event types of shorter duration. Again, the trend indicates that the observed sample may not represent a true picture of the sequence factor in the curriculum. It reinforces the assumption that a review and reinforcement factor is present in the data and may have to be considered in the estimating of parameters for the variables in the model. With all of these observations noted, it was still assumed that the observed samples selected were adequate for judging the descriptive ability of the model. However, these observations must be taken into consideration in observing the graphs of the simulated and observed distributions.

Runs of the model were made using parameters from the "participating" samples. Each computer run consisted of ten samples of the number of students on which each of the observed samples was based (thirty-six for Level B; fifty-six for Level C; *etc*.). The output from these computer runs were combined into frequence distributions of the percent of students falling into each of several time intervals and compared to the appropriate observed sample. Figures 4.3 and 4.4 display the results for Levels B and D.

FIGURE 4.3 : Observed and Simulated Distributions for Level B Based on 1967-68 Data for Time to Complete a Level Within One School Year.

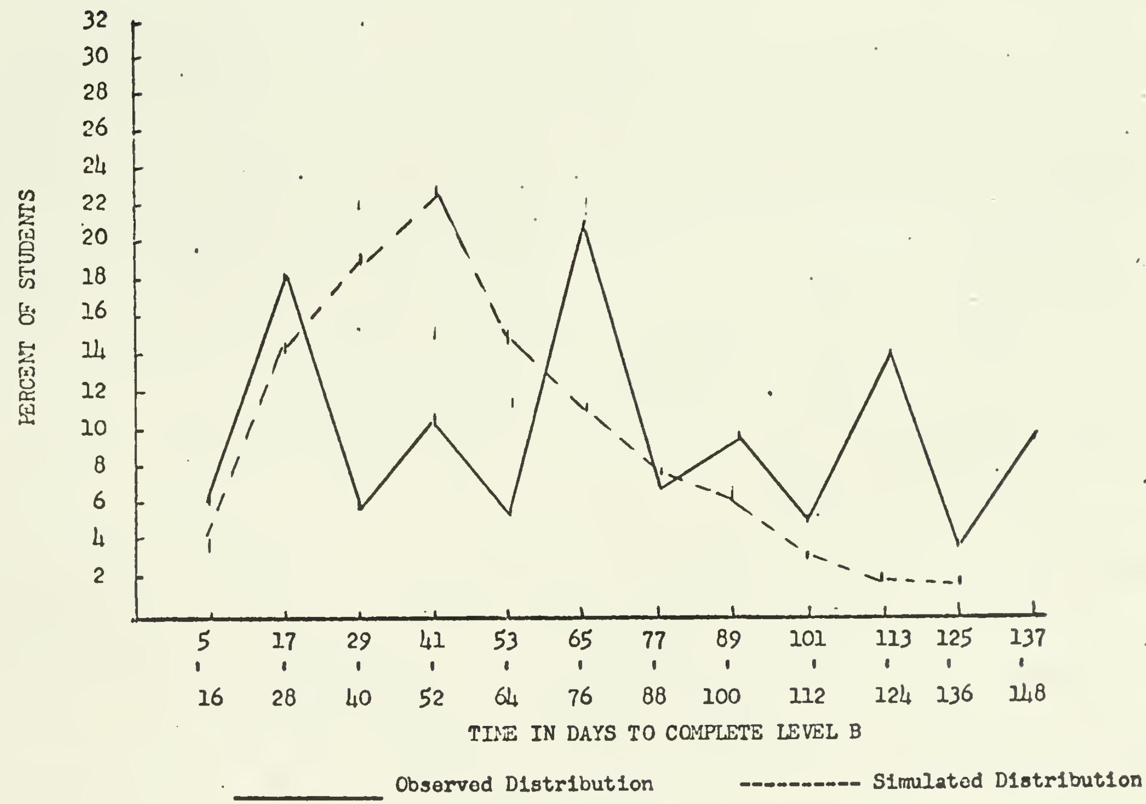
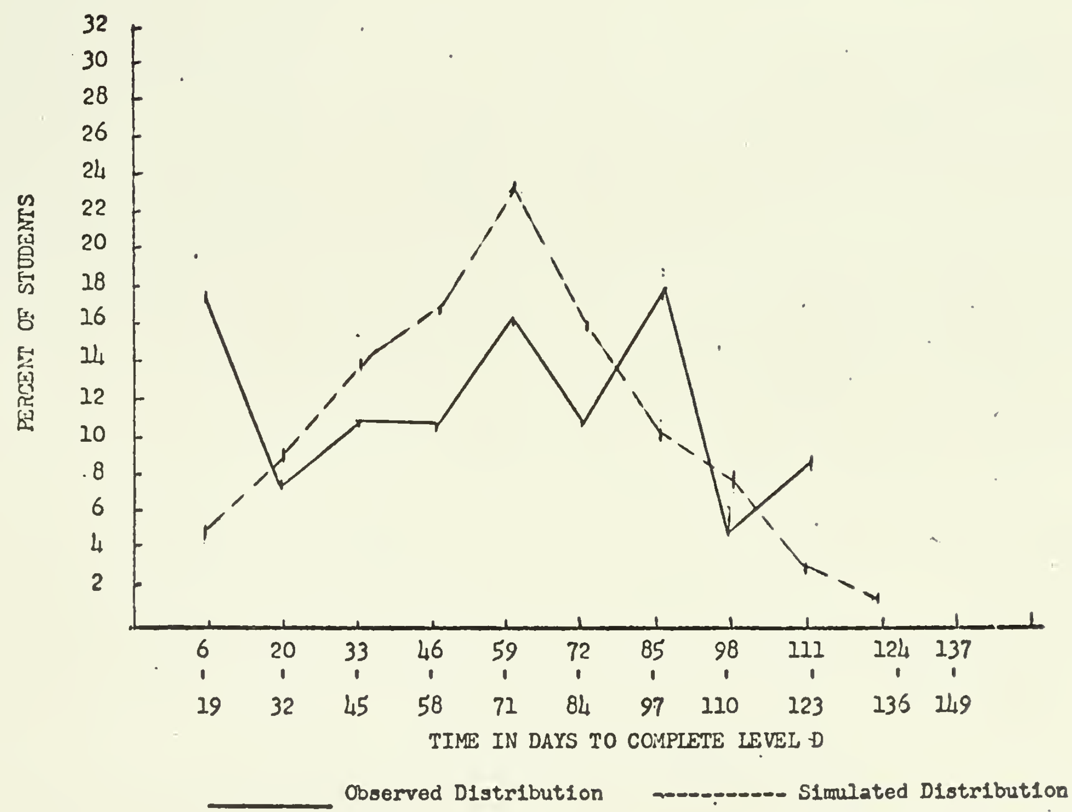


Figure 4.4: Observed and Simulated Distributions for Level D Based on 1967-68 Data for Time to Complete a Level Within One School Year.



Several observations can be made about the distributions in Figures 4.3 and 4.4. The simulated sample produces completion times slightly higher than those of the observed sample. However, it can be observed that the two distributions are similar in range of time to completion. The range in the Level D observed sample (eleven days to 120 days) was 110 days; the range of the simulated sample (eighteen days to 133 days) was 115 days. The range of the Level B observed sample (eight days to 138 days) was 131 days; the range of the simulated sample (eleven days to 127 days) was 117 days. While this statistic cannot be used in making great conclusions about the model's ability to describe the system, it does show that the sampling procedures used in the model have the ability to produce ranges of times to completion that are realistic.

It can also be observed that the simulated distribution has a positive skew; part of which could be due to the review factor mentioned earlier that appears to be present in the observed sample. Runs of the model were made using parameters from the observed (completing) sample in order to produce a simulated sample that could be compared to the observed sample. Even with parameters so exactly estimated, the positive skew is reduced but is still present. This question of skewness is analyzed later in this study in regard to times before and after a post-test as an improvement to the model.

It can also be observed from the distributions in Figures 4.3 and 4.4 that the simulated distribution produces a smooth curve, and one that approaches a normal curve. This is to be expected in a model of this type that samples from a probability distribution and sums a series of random variables to get a total time. This is simply a graphic representation of the central limit theorem. It can be hypothesized that in the real world certain students would always require low service times and have high probabilities of passing a pretest or a post-test and that certain students would always require high service times and have low probabilities of passing a pretest or a post-test. A model that deals with this hypothesis would use student attributes as a variable; however, further research is being conducted to identify relationships between time in the system and student attributes. This research has as its ultimate aim the construction of a guidance subroutine to be added to the EDSIM 1 model.

Statistical analysis of the relationships between the simulated and observed distributions was made using the Kolmogorov-Smirnoff Two Sample Test. This test provides a measure of the agreement between two cumulative distributions, to answer the hypothesis that sample distributions were drawn from the same population. Since the concern is for difference regardless of direction, a two-tailed test was used. This test is sensitive to any kind of difference in the distributions: central tendency, dispersion, skewness, and range.

If the two samples are representative of the same population then the cumulative distributions of both samples may be expected to be fairly close to each other. While a definite level of significance can be chosen to determine whether the distributions are significantly different, the important aspect of these tests was to determine how large a difference did exist. It could be that a user of the model would accept a large difference in the two distributions as long as there is some relationship since the only other choice of determining completion times could be an intuitive guess. In that case, any information garnered from a simulation model could be better than no information at all.

However, in applying the test in this study it was felt that a difference above the significance level of .05 would be realistic in determining whether the model had descriptive validity. When parameters had been estimated as carefully as these had from historical data, any greater difference would leave doubt as to the possibility of using the model to predict time in a level if it could not describe the system with more accuracy.

In each test of the simulated distributions, the following hypotheses were tested:

H_0: the samples are from populations with the same distributions or are from the same population.

H_1: the samples are from populations with different distributions.

Significance level: as stated earlier, a significance level of .05 was selected.

Tables 4.4, 4.5 and 4.6 show the pertinent statistical data for these tests for Levels B, D and E. The intervals that are used for the distributions are based on the output histogram produced by the EDSIM 1 model since this would be the intervals considered by a user of the model in judging time to completion of a level.

The tests indicate that Levels B and D have differences (.2471 and .2014) that are significant at the .05 level. Level E shows a

Table 4.4

Data Cast for Kolmogorov-Smirnov Two Sample Test for Comparison of Observed Data with Simulated Data Using Service Times Based on Time in the Unit Before a Post Test - Level E

	TIME IN DAYS TO COMPLETE LEVEL E											
	1 16	17 33	34 50	51 67	68 84	85 101	102 118	119 135	136 152	153 169	170 186	187 203
Frequency count of observed data S_{38}	1	2	3	1	11	6	6	4	2	1	1	0
Frequency count of simulated sample S_{38}	0	0	8	27	57	95	76	41	41	19	8	8
Cumulative Distribution $S_{38}(X)$	$\frac{1}{38}$	$\frac{3}{38}$	$\frac{6}{38}$	$\frac{7}{38}$	$\frac{18}{38}$	$\frac{24}{38}$	$\frac{30}{38}$	$\frac{34}{38}$	$\frac{36}{38}$	$\frac{37}{38}$	$\frac{38}{38}$	$\frac{38}{38}$
Cumulative Distribution $S_{380}(X)$	$\frac{0}{380}$	$\frac{0}{380}$	$\frac{8}{380}$	$\frac{35}{380}$	$\frac{92}{380}$	$\frac{187}{380}$	$\frac{263}{380}$	$\frac{304}{380}$	$\frac{345}{380}$	$\frac{364}{380}$	$\frac{372}{380}$	$\frac{380}{380}$
$S_{38}(X) - S_{380}(X)$	$\frac{10}{380}$	$\frac{30}{380}$	$\frac{52}{380}$	$\frac{35}{380}$	$\frac{88}{380}$	$\frac{53}{380}$	$\frac{37}{380}$	$\frac{36}{380}$	$\frac{15}{380}$	$\frac{6}{380}$	$\frac{8}{380}$	0

$D = \text{Maximum } S_{N1}(X) - S_{N2}(X) = \frac{88}{380} = .2301$

Level of Significance at .01 = 1.63(.17) = .2771 at .05 = 1.36(.17) = .2312

Table 4.5

Data Cast for Kolmogorov-Smirnov Two Sample Test for Comparison of Level D Observed Data with Simulated Data Using One Service Time Distribution

	TIME IN DAYS TO COMPLETE LEVEL D									
	6 - 19	20 - 32	33 - 45	46 - 58	59 - 71	72 - 84	85 - 97	98 - 110	111 - 123	124 - 136
Frequency of Observed Sample S_{70}	13	6	9	11	12	9	3	3	4	
Frequency of Simulated Sample S_{700}	7	42	92	112	147	121	91	56	19	14
Cumulative Distribution of Observed $S_{70}(X)$	$\frac{13}{70}$	$\frac{19}{70}$	$\frac{28}{70}$	$\frac{39}{70}$	$\frac{51}{70}$	$\frac{60}{70}$	$\frac{63}{70}$	$\frac{66}{70}$	$\frac{70}{70}$	$\frac{70}{70}$
Cumulative Distribution of $S_{700}(X)$	$\frac{7}{700}$	$\frac{49}{700}$	$\frac{141}{700}$	$\frac{253}{700}$	$\frac{400}{700}$	$\frac{521}{700}$	$\frac{611}{700}$	$\frac{667}{700}$	$\frac{686}{700}$	$\frac{700}{700}$
$S_{70}(X) - S_{700}(X)$	$\frac{123}{700}$	$\frac{141}{700}$	$\frac{139}{700}$	$\frac{137}{700}$	$\frac{110}{700}$	$\frac{70}{700}$	$\frac{19}{700}$	$\frac{7}{700}$	$\frac{14}{700}$	—

$$D = \text{Maximum} \quad S_{70}(X) - S_{700}(X) = \frac{141}{700} = .2014$$

Level of Significance at .05 = 1.36 (.124) = .1566

.05 = 1.36 (.124) = .1566

Table 4.6

Data Cast for Kolmogorov-Smirnoff Two Sample Test for Comparison of Level B Observed Data with Simulated Data Using One Service Time Distribution

	TIME IN DAYS TO COMPLETE LEVEL B											
	5 - 16	17 - 28	28 - 40	41 - 52	53 - 64	65 - 76	77 - 88	89 - 100	101 - 112	113 - 124	125 - 136	137 - 148
Frequency Count of Observed Data S_{36}	2	6	2	4	2	7	2	3	2	3	1	2
Frequency Count of Simulated Sample S_{360}	14	53	62	72	48	40	36	19	9	5	2	0
Cumulative Distribution S_{38} (X)	$\frac{2}{36}$	$\frac{8}{36}$	$\frac{10}{36}$	$\frac{14}{36}$	$\frac{16}{36}$	$\frac{23}{36}$	$\frac{25}{36}$	$\frac{28}{36}$	$\frac{30}{36}$	$\frac{33}{36}$	$\frac{34}{36}$	$\frac{36}{36}$
Cumulative Distribution of Simulated Sample S_{360}	$\frac{14}{360}$	$\frac{67}{360}$	$\frac{129}{360}$	$\frac{201}{360}$	$\frac{249}{360}$	$\frac{289}{360}$	$\frac{325}{360}$	$\frac{344}{360}$	$\frac{353}{360}$	$\frac{358}{360}$	$\frac{360}{360}$	$\frac{360}{360}$
S_{38} (X) - S_{380} (X)	$\frac{6}{360}$	$\frac{13}{360}$	$\frac{29}{360}$	$\frac{61}{360}$	$\frac{89}{360}$	$\frac{59}{360}$	$\frac{75}{360}$	$\frac{64}{360}$	$\frac{53}{360}$	$\frac{28}{360}$	$\frac{20}{360}$	—

$$D = \text{Maximum} \quad S_{38}(X) - S_{380}(X) = \frac{89}{360} = .2471$$

Level of Significance at .05 = 1.36 (.175) = .2380

difference of .2301 which is not significant at the .05 level but does represent a large difference in the distributions. These findings lead to the conclusion that some improvement would be necessary before the model could be expected to accurately predict the system operation in terms of time required to complete a level.

Improvements to Increase Descriptive Ability of Model

One of the original assumptions in the building of the EDSIM 1 model was that combining service times before and after a post-test would produce a service time distribution that could be sampled each time a pretest was failed as well as when a post-test was failed. Observation of the data indicated that this assumption could in fact be erroneous and could be causing the positive skew in the distributions.

Table 4.7 shows a sample of the service time distributions for Level E when separate distributions are used for time in a unit based on time distribution data before a post-test and time after a post-test. From the data displayed in Table 4.7, it is evident that service times after a post-test tend to have a narrower range and higher probabilities associated with event types of shorter duration than the before post-test service times. These observations lead to the hypothesis that if two service times distributions (one based on time before a post-test and one based on time after a post-test) could be input into the model, its descriptive ability could be improved.

The model was then changed to permit two service time distributions to be used. Runs of the revised model were made to test the above

Table 4.7

Comparison of Eight Units of Level E Showing Service Time Distributions
Both Before and After a Post Test

	UNITS															
	1		2		3		4		5		6		7		8	
	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A
ET 1	.00	.19	.00	.34		.12	.07	.18	.11	.05	.11	.30	.05	.20	.11	.12
ET 2	.06	.07	.00	.33		.11		.08	.07	.05		.36	.13	.25	.11	.11
ET 3	.00	.07	.11		.50	.11	.06	.08	.04	.05	.04		.05	.25	.04	.11
ET 4	.03	.15	.00	.33		.11	.13	.08	.03	.05	.04	.07	.05		.03	.11
ET 5	.00	.07	.22			.11	.06	.08		.05	.04		.13		.03	.11
ET 6	.15	.07	.33		.17	.11	.13	.18		.05				.05	.11	.11
ET 7	.11		.12			.11	.06			.05	.04	.07			.11	.11
ET 8	.06	.11	.11			.11	.06	.08	.03	.05			.13			.11
ET 9	.06	.07				.11	.06	.08		.05	.08		.13	.12	.11	.11
ET10	.17	.04					.06			.05	.07	.10	.09		.07	
ET11	.09	.04					.06	.08	.14	.05	.04		.05		.03	
ET12	.06				.33		.06		.03	.05	.11			.13	.11	
ET13	.06	.04	.11				.06	.08	.11	.05	.04	.10	.09			
ET14	.03	.04								.05						
ET15	.00									.05	.04				.07	
ET16	.00	.04							.11	.05	.04		.05		.03	
ET17	.06						.06			.05			.05			
ET18	.03						.07		.11	.05	.07					
ET19										.05	.08				.04	
ET20									.04	.05						
ET21									.04		.04					
ET22											.04					
ET23											.04					
ET24	.03								.07							
ET25											.04					

B = Service Times Before a Post Test .

A = Service Times After a Post Test

hypothesis. All conditions except service time distributions were held constant with runs made of the original model including the same sequence of random numbers.

The results of the run for Level E are shown in Figure 4.5. The Kolmogorov-Smirnoff Two Sample Test was applied to the data. Table 4.8 is the pertinent data for this statistical test for Level E. It can be observed that the difference is reduced from .2013 for one service time distribution (Table 4.4) to .087 for the run with two service time distributions. Subsequent runs and tests were made for the other levels. In the Level B sample D = 40/360 = .1111 for the simulated distribution from the improved model as compared to the .2471 (Table 4.6) for the original model. In the Level D sample D = 72/700 = .1078 as compared to the .2014 (Table 4.5) for the original model. These reductions in difference show that the model has been improved in its descriptive ability. The critical value of D for all samples was less than the required value for rejection at the .10 level of significance (Siegel; Table M, p. 279; 1956).

The EDSIM 1 model also produces output in regard to the average time spent in a unit. Table 4.9 shows the actual versus the simulated time results for data in Level E. It can be observed from this data that even though the model produces higher simulated total times to completion of a level it consistently produces average times lower than those from the observed data.

Investigation of this aspect of the model revealed that average time for a unit being produced in the model were concerned with average

FIGURE 4.5: Comparison of Observed Data to Simulated Distributions for Sample of 38 Students That Completed Level E in 1967-68.

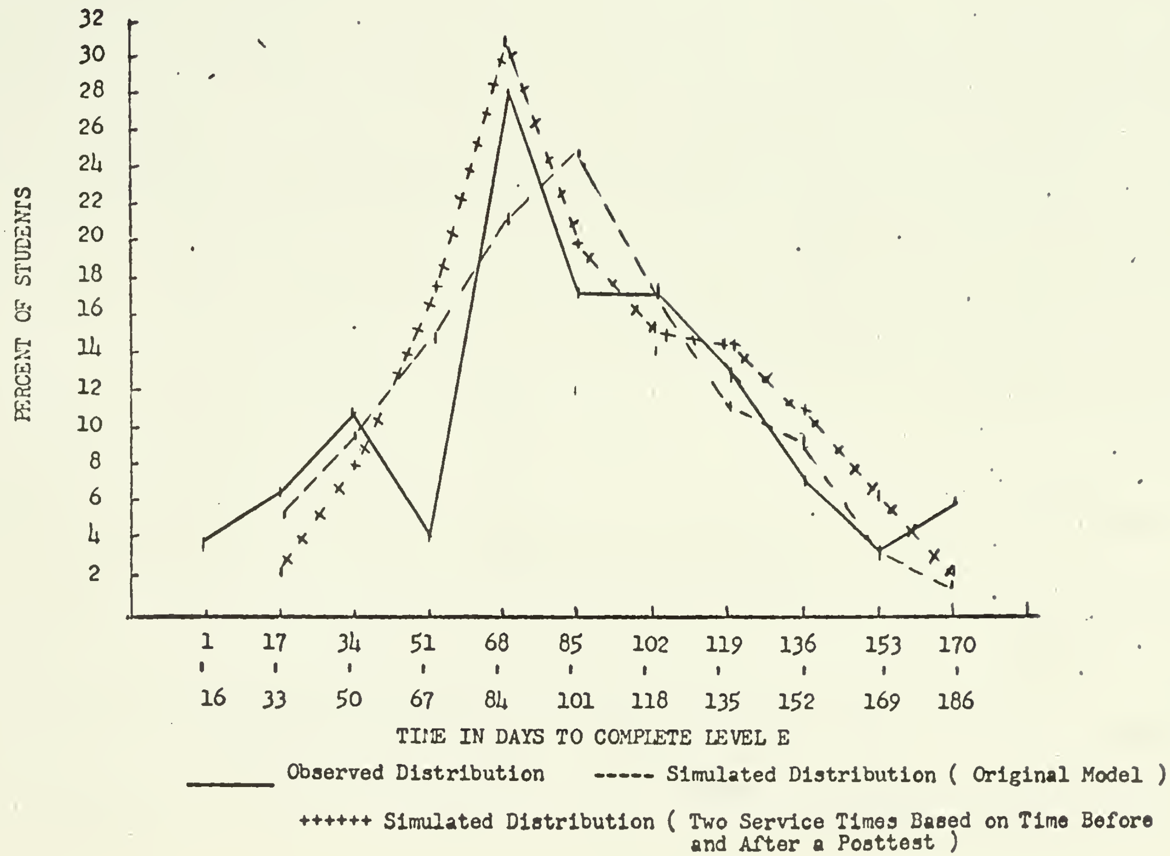


Table 4.8

Data Cast for Kolmogorov-Smirnov Two Sample Test for Comparison of Observed Data With Simulated Data Using Before and After Post Test Service Time Distributions

	TIME IN DAYS TO COMPLETE LEVEL E										
	1 - 16	17 - 33	35 - 50	51 - 67	68 - 84	85 - 101	102 - 118	119 - 135	136 - 152	153 - 169	170 - 186
Frequency count of observed data - S_{38}	1	2	3	1	11	6	6	4	2	1	1
Frequency count of simulated distribution - S_{380}	0	8	19	56	98	63	47	40	30	15	4
Cumulative Distribution - $S_{38}(X)$	$\frac{1}{38}$	$\frac{3}{38}$	$\frac{6}{38}$	$\frac{7}{38}$	$\frac{18}{38}$	$\frac{24}{38}$	$\frac{30}{38}$	$\frac{34}{38}$	$\frac{36}{38}$	$\frac{37}{38}$	$\frac{38}{38}$
Cumulative Distribution - $S_{380}(X)$	$\frac{0}{380}$	$\frac{8}{380}$	$\frac{27}{380}$	$\frac{83}{380}$	$\frac{181}{380}$	$\frac{244}{380}$	$\frac{291}{380}$	$\frac{331}{380}$	$\frac{361}{380}$	$\frac{376}{380}$	$\frac{380}{380}$
$S_{38}(X) - S_{380}(X)$	$\frac{10}{380}$	$\frac{22}{380}$	$\frac{33}{380}$	$\frac{13}{380}$	$\frac{1}{380}$	$\frac{4}{380}$	$\frac{9}{380}$	$\frac{9}{380}$	$\frac{1}{380}$	$\frac{6}{380}$	0

$$D = \text{Maximum} \quad S_{N1}(X) - S_{N2}(X) = \frac{33}{380} = .087$$

$$\text{Level of Significance at } .10 = 1.22 \quad \frac{N1 + N2}{N1N2} = .2074$$

Table 4.9

Comparison of Average Unit Times Produced by Model
to Those of Observed Data

LEVEL E UNIT	SAMPLE OF 38 COMPLETING E OBSERVED DATA	PARAMETERS FROM SAMPLE OF 38 COMPLETING E SIMULATED DATA*	PARAMETERS FROM SAMPLE OF 103 PARTICIPATING IN E SIMULATED DATA*	SAMPLE OF 103 PARTICIPATING IN E OBSERVED DATA
1	15.3	16.8	17.3	18.6
2	7.1	6.6	4.0	8.4
3	8.6	4.5	7.0	5.8
4	13.1	9.7	4.6	12.1
5	17.5	15.3	14.6	18.1
6	14.1	15.2	12.8	13.3
7	11.1	9.1	5.7	6.3
8	10.6	10.2	9.2	10.1
9	4.4	3.9	3.3	3.1
10	13.9	9.8	8.6	11.1
11	14.7	10.5	9.2	10.8
12	11.5	11.2	10.3	12.1

*Average produced from run of ten samples of thirty-eight students time for all students in the system (level) and not for only those that participated in a unit. If 100 students were used as input into the model, this number becomes the denominator for finding the average time in a unit even though twenty-five have pretested out of the unit. This is a reasonable method for considering average time in a unit. However, for a system such as IPI it was felt that average time in a unit should reflect the average time spent in a unit by those students that actually participated in the unit. Arguments can be made for each approach. The most important aspect, however, is that the user of the model be aware of the meaning attached to the synthetic output from the model.

The model was changed to reflect average time for those students participating in a unit. Table 4.10 shows the results of the output of average time in a unit after the change was made in the model.

It can be observed from Table 4.10 that simulated averages are very close to observed averages in unit completion times. What use can be made of these averages is questionable in a system such as IPI. They could be used to determine if certain students are spending excessively long periods of time in a unit. However, it must be noted that most of the observed distributions tend to be bimodal. This observation must be considered in using the average unit times as a part of the system description.

Table 4.10

Comparison of Average Unit Times Produced by Model to Those of Observed Data After Change to Consider Average Time of Only the Students That Did Not Pretest Out of Unit

LEVEL E UNIT	SAMPLE OF 38 COMPLETING E OBSERVED DATA	PARAMETERS FROM SAMPLE OF 38 COMPLETING E SIMULATED DATA*	PARAMETERS FROM SAMPLE OF 103 PARTICIPATING IN E SIMULATED DATA*	SAMPLE OF 103 PARTICIPATING IN E OBSERVED DATA
1	15.3	17.1	19.3	18.6
2	7.1	8.0	8.5	8.4
3	8.6	8.9	8.1	5.8
4	13.1	13.9	13.8	12.1
5	17.5	19.1	20.1	18.1
6	14.1	15.2	14.1	13.3
7	11.1	12.3	6.8	6.3
8	10.6	11.6	11.3	10.1
9	4.4	4.3	3.9	3.1
10	13.9	13.8	12.3	11.1
11	14.7	16.1	11.2	10.8
12	11.5	11.9	13.3	12.1

*Average produced from run of ten samples of thirty-eight students

Description Across Years

The next tests to be conducted in the descriptive phase were tests to assess the ability of the model to describe time to completion in a level, when the level had been completed over a two year span. Many students began a level in one school year but did not complete it until the middle or end of the next school year. The question to be answered was whether a variable labeled "summer" would have to be included if the model were to describe time across years accurately.

Careful observation of the data for 1967-1968 and 1968-1969 indicated that summer seemed to have little effect on time needed to complete a level. It could be observed that students who had spent little time in the numeration unit might spend much longer in the addition unit than a student who had spent a long time in the numeration unit. This type of trend suggested an additive feature that was present and should be considered in estimating parameters. This review or reinforcement factor that appeared to be present in the data did not appear to be dependent upon summer.

The data for those students that had participated in a level during 1967-1968 and 1968-1969 were used in estimating the parameters for these tests of the model. Data for students who had participated in a specific level in 1967-1968 were converted into input probabilities for service time distributions. The data for students who had participated in a specific level in 1968-1969 were also converted into input probabilities for service time distributions. Since some of the students in the 1968-1969 school year were to be used as a sample in determining the predictive ability of the model, the data for these students were not used in the parameter estimation for these tests. The two sets of input probabilities were then added and an average taken to get unit service time distributions. These averages were not weighted since preliminary examination of weighted averages showed almost no effect on the final unit distributions. The participating samples for each year were very nearly equal. The pretest and post-test probabilities for 1967-1968 and 1968-1969 were also averaged for input into the model.

Runs of the model were made using the above parameters and compared to samples of students that had completed a specific level either within a year or across two years. For Level D, this included a sample of 109; for Level B, fifty-two; for Level C, seventy-eight; and for Level E, fifty-seven. Data for Level E proved very difficult to determine since some students repeated material in this level for a three year period.

Using the estimated parameters established across years, runs of the model were made for all levels. Figure 4.6 and 4.7 show the results for Levels B and D. It was hypothesized that the model should be able to describe the system across years with the information contained in the estimated parameters. It can be observed that the real distribution of Level B tends to be bimodal with high frequencies towards the ends of the distribution. The simulated distribution is also bimodal but the high frequencies have moved more towards the center of the distribution. It can be observed that although the observed distribution of Level D has a wider range than the simulated distributions, the two distributions

FIGURE 4.6: Comparison of Level B Observed and Simulated Distributions Across Years (1967-68 and 1968-69).

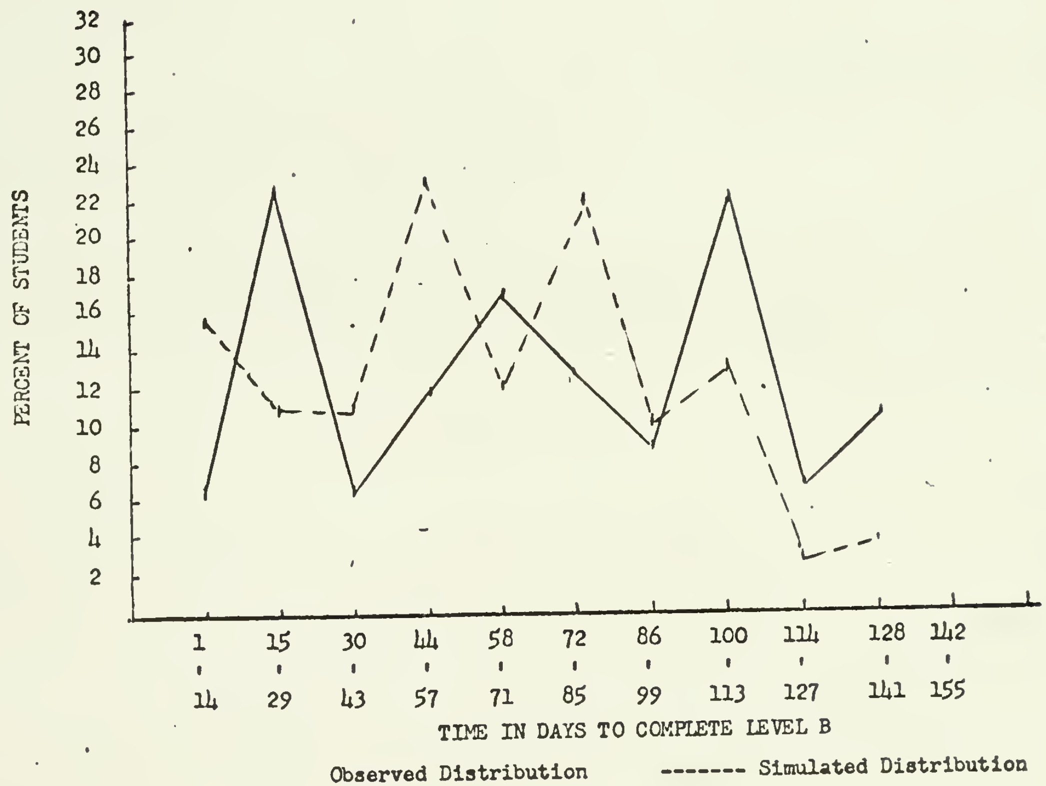
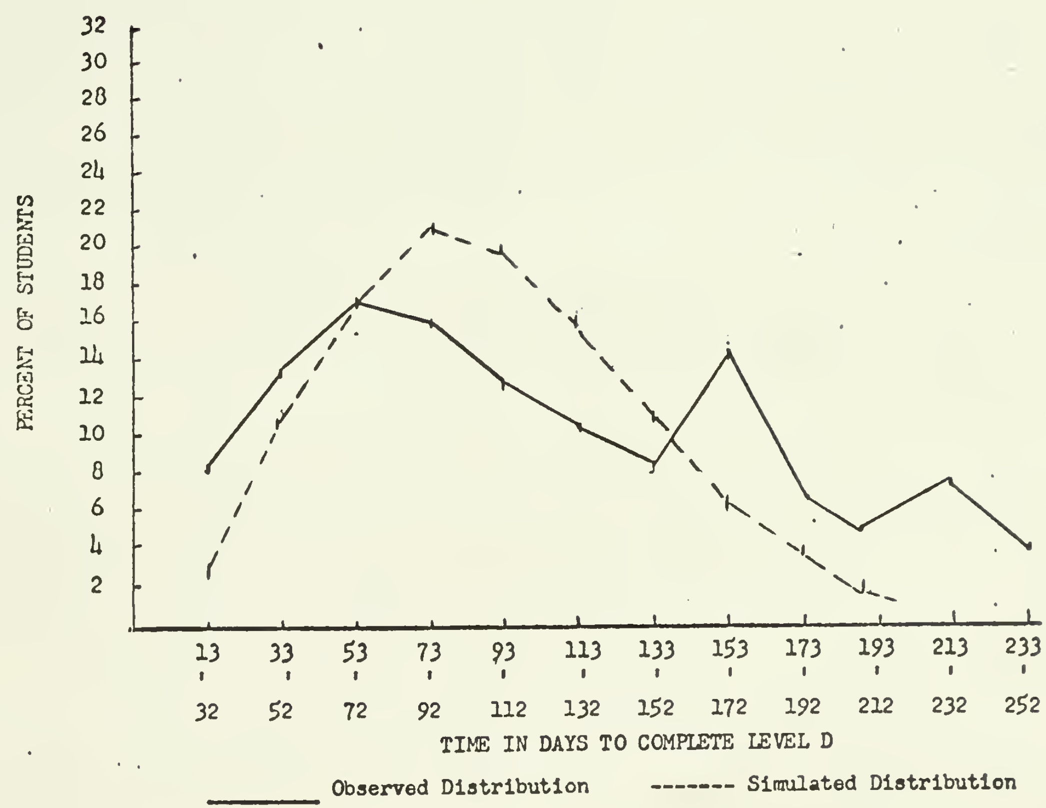


FIGURE 4.7: Comparison of Observed and Simulated Distributions for Level D Across Years (1967-68 and 1968-69).



do have similar characteristics. In both cases, the simulated distributions show major differences in the extreme tails of the distribution. This lost of information in the tails of the distribution is to be expected in a model that does not include student attributes.

In order to check the statistical difference of these distributions, the Kolmogorov-Smirnoff Two Sample Test that had been used to determine differences within years was applied. Table 4.11 and 4.12 show the statistical information for these tests for Levels B and D. The differences that exist are greater than the differences for the samples describing within years but are less than the required difference to reject the hypothesis that the samples tested are from populations with the same distributions (.05 level of significance).

These tests indicate that the model is capable of describing across years as well as within years when parameters are carefully estimated. If very much system change occurred over these two years, it was compensated for by the careful estimation of parameters. The problem of system change is a very important aspect to consider when using the model for prediction purposes. System changes would have to be considered since marked changes in sequence or content would need to be considered in adjusting parameters.

Table 4.11

Data Cast for Kolmogorov-Smirnoff Two Sample Test for Comparison of Level B Observed Sample and Simulated Sample Across Years (1967-68 & 1968-69)

	TIME IN DAYS TO COMPLETE LEVEL B									
	1-14	15-29	30-43	44-57	58-71	72-85	86-99	100-113	114-127	128-141
Frequency of Observed Sample S_{52}	3	9	3	5	7	5	4	9	3	4
Frequency of Simulated Sample S_{520}	78	52	52	88	52	78	36	52	16	16
Cumulative Distribution of Observed Sample S_{52} (X)	$\frac{3}{52}$	$\frac{12}{52}$	$\frac{15}{52}$	$\frac{20}{52}$	$\frac{27}{52}$	$\frac{32}{52}$	$\frac{36}{52}$	$\frac{45}{52}$	$\frac{48}{52}$	$\frac{52}{52}$
Cumulative Distribution of Simulated Sample S_{520} (X)	$\frac{78}{520}$	$\frac{130}{520}$	$\frac{182}{520}$	$\frac{270}{520}$	$\frac{322}{520}$	$\frac{400}{520}$	$\frac{436}{520}$	$\frac{488}{520}$	$\frac{504}{520}$	$\frac{520}{520}$
S_{52} (X) - S_{520} (X)	$\frac{48}{520}$	$\frac{10}{520}$	$\frac{32}{520}$	$\frac{70}{520}$	$\frac{52}{520}$	$\frac{80}{520}$	$\frac{76}{520}$	$\frac{38}{520}$	$\frac{24}{520}$	0

$$D = \text{Maximum} \quad S_{52}(X) - S_{520}(X) = \frac{80}{520} = .1538$$

Level of Significance at .05 = 1.36 (.145) = .1538

Table 4.12

Data Cast for Kolmogorov-Smirnoff Two Sample Test for Comparison of Level D Observed Sample and Simulated Sample Across Years (1967-68 & 1968-69)

	TIME IN DAYS TO COMPLETE LEVEL D											
	13 - 32	33 - 52	53 - 72	73 - 92	93 - 112	113 - 132	133 - 152	153 - 172	173 - 192	193 - 212	213 - 232	233 - 252
Frequency of Observed Sample S_{109}	9	14	17	15	14	11	7	8	6	3	4	1
Frequency of Simulated Sample S_{1090}	58	98	165	232	220	159	83	48	14	13	0	0
Cumulative Distribution of Observed Sample S_{109} (X)	$\frac{9}{109}$	$\frac{23}{109}$	$\frac{40}{109}$	$\frac{55}{109}$	$\frac{69}{109}$	$\frac{80}{109}$	$\frac{87}{109}$	$\frac{95}{109}$	$\frac{101}{109}$	$\frac{104}{109}$	$\frac{108}{109}$	$\frac{109}{109}$
Cumulative Distribution of Simulated Sample S_{1090} (X)	$\frac{58}{1090}$	$\frac{156}{1090}$	$\frac{321}{1090}$	$\frac{553}{1090}$	$\frac{713}{1090}$	$\frac{932}{1090}$	$\frac{1015}{1090}$	$\frac{1063}{1090}$	$\frac{1077}{1090}$	$\frac{1090}{1090}$	$\frac{1090}{1090}$	$\frac{1090}{1090}$
S_{109} (X) - S_{1090} (X)	$\frac{32}{1090}$	$\frac{74}{1090}$	$\frac{79}{1090}$	$\frac{3}{1090}$	$\frac{23}{1090}$	$\frac{132}{1090}$	$\frac{145}{1090}$	$\frac{113}{1090}$	$\frac{67}{1090}$	$\frac{50}{1090}$	$\frac{10}{1090}$	0

$$D = \text{Maximum} \quad S_{109}(X) - S_{1090}(X) = \frac{145}{1090} = .1330$$

Level of Significance at .05 = 1.36 (.101) = .1373

Prediction Phase

The parameters established for 1967-1968 and 1968-1969 were used to test the model's ability to predict time to completion of a specific level for selected samples of students for 1968-1969 and 1969-1970.

Runs of the model were made using the 1967-1968 and 1968-1969 parameters and compared to samples of students that had completed a level either during 1969-1970 or over a two year period 1968-1969 through 1969-1970. Observation of the data showed that Unit 13 (Special Topics) began to appear in the data for 1969-1970. This was the first year that data concerning this unit had been available. Information concerning the pretest and post-test probabilities and service time distributions were collected for this unit and parameter estimations were made. The range of time spent in this unit was very narrow and it appeared that its addition to the system parameters would not extend completion time very much. This additional information was input into the model, however.

Observation of the data also revealed that the number of units required to complete a level had not changed except for the addition of Unit 13. It still required the same number of units to complete Levels B, C, D, and E that had been required in other years. Level B required eight units, Level C required eleven units, etc., as described in Table 3.1. Any change that had occurred in the skills within these units would be difficult to assess in adjusting unit parameters since information at this level had not been the objective of this study.

Examination of the data had indicated that stability in parameters from year to year would be higher in the lower levels (B and C). Rank correlations were performed on pretest and post-test probabilities for 1967-1968 and 1968-1969 for Levels B, C, D and E. Correlation of pretest data showed that Level B probabilities for 1967-1968 correlated .88 with the probabilities determined for 1968-1969. The same information correlated for other levels produced coefficients of .82 for Level C; .68 for Level D; and .28 for Level E. This indicates that the longer students are in the system the more difficult stability of parameter estimation will be. Attempts to correlate service time distributions across years was made. No correlations that were significant from zero could be found at any level for any unit. Since one of the most important variables in the model was service time distributions, this indicated that the model's prediction ability would probably be poor.

Figures 4.8 and 4.9 display the results of the prediction runs made for Levels B and D. Parameters had been adjusted as accurately as possible to compensate for system changes in regard to curriculum content and sequence. It would be expected that different approaches are always being tried in a system such as IPI. That is, one group may be used to experiment with a new approach to the system operation. This type of change that is not operational over the entire system is difficult to account for in parameter estimation in a model as simple as EDSIM 1. It could also be hypothesized that a system such as IPI would cease to be worthwhile once it is predictable or becomes stable and inflexible.

FIGURE 4.8: Comparison of Predicted (Simulated) Distribution to Observed Distribution (Sample from 1968-69 and 1969-70 Data).

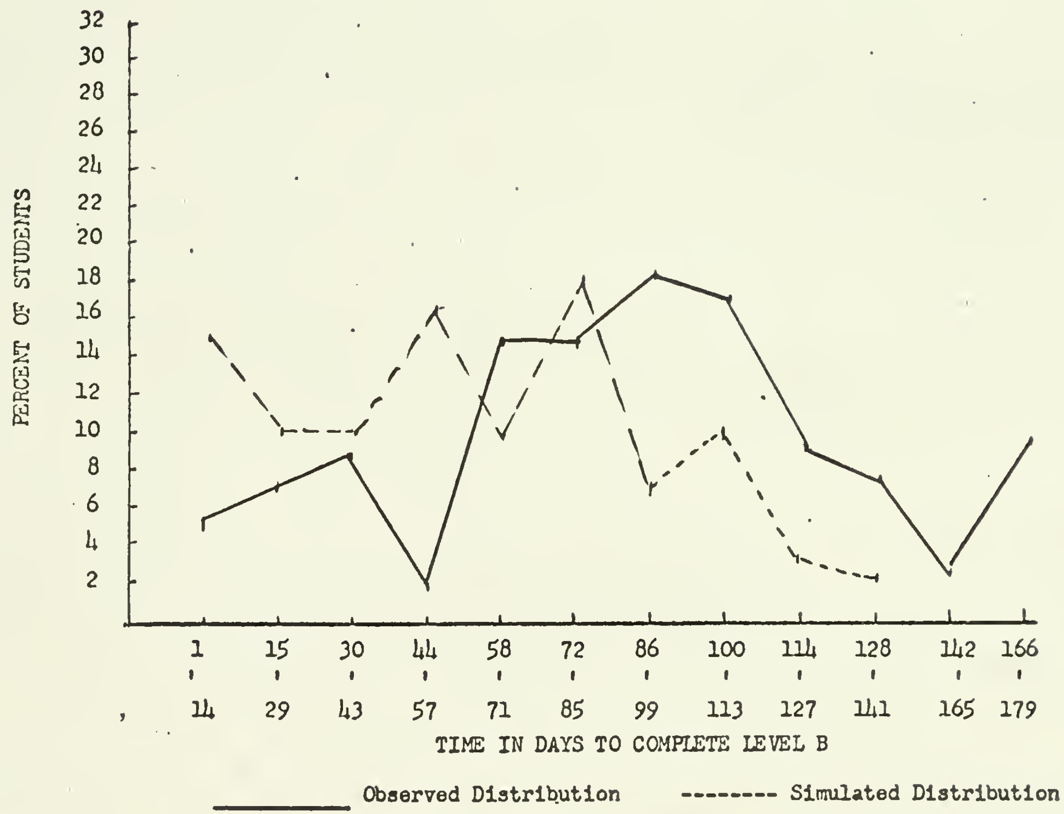
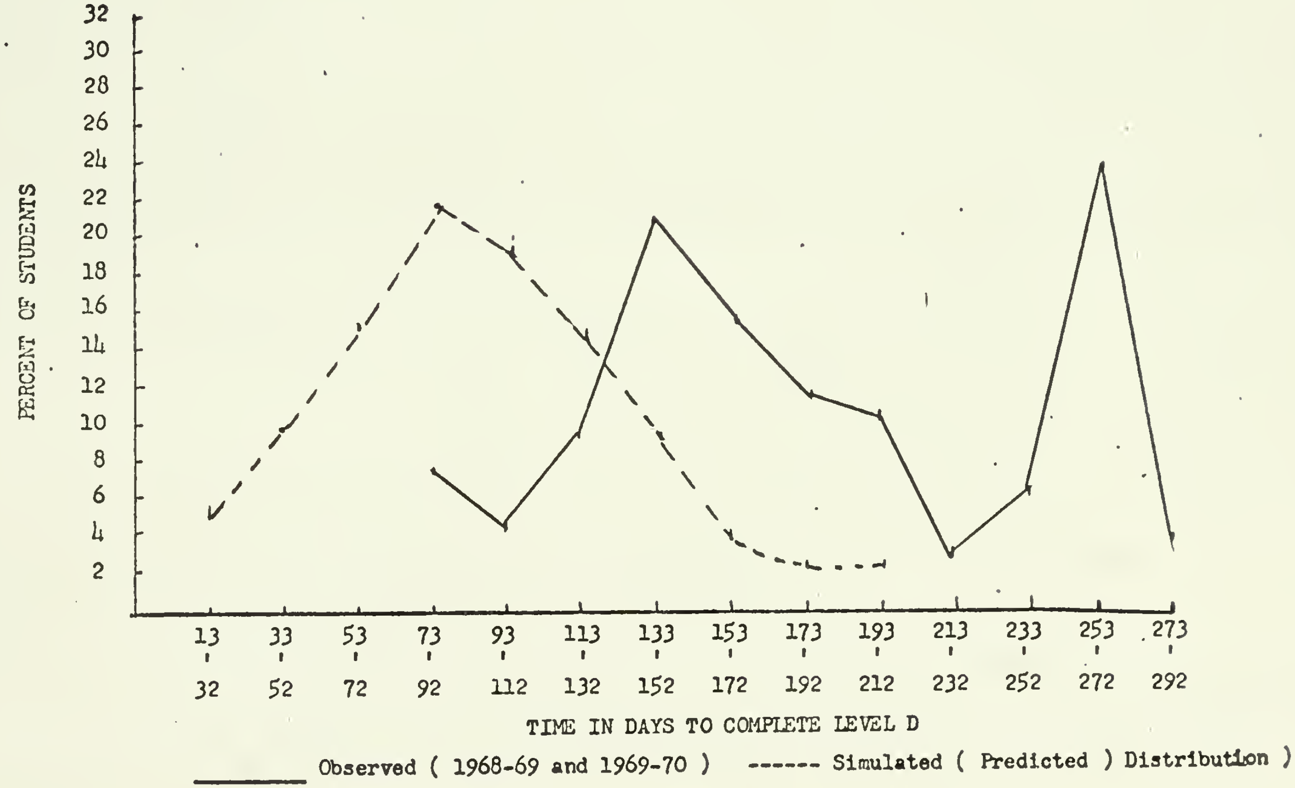


FIGURE 4.9: Comparison of Predicted (Simulated) Distribution to Observed Distribution (Sample from 1968-69 and 1969-70 Data).



Figures 4.8 and 4.9 show that there is more relationship between the Level B simulated and real distributions than can be found in the Level D distributions. In both cases, however, the simulated distributions have narrower ranges and higher frequencies in the lower time intervals than do the observed distributions. Application of the Kolmogorov-Smirnoff Two Sample Test shows a difference of 183/600 = .3500 for the Level B Sample and 495/990 = .5000 for Level D. Level B produced the smallest statistical difference and Level E produced the largest difference (.5268). This provides statistical proof that the hypothesis that the higher the level and thus the longer students remain in the system, the harder it will be to make accurate parameter estimations for accurate predictions. All differences produced in the statistical tests were above the .001 rejection level. This indicates no statistical evidence to support the hypothesis that the samples are from populations with the same distributions. Statistically, this means that no relationships can be shown between the simulated and observed distributions. However, when graphed as in Figures 4.8 and 4.9, these distributions can indicate relationships and information to a potential user of the model.

It can be observed in both Figure 4.8 and Figure 4.9 that the observed and simulated distributions have comparable ranges. In the case of Level D, there is a definite trend toward longer times in the level. The shortest amount of time any student spent in the observed sample was seventy-five days compared to fifteen days for the simulated sample. This is a good indication that a system change has occurred. It does not appear as a change in the original units to be completed in

Table 4.13

Data Cast for Kolmogorov-Smirnoff Two Sample Test for Comparison of Observed Sample and Simulated (Predicted) Sample for Level B (1968-69 & 1969-70)

	TIME IN DAYS TO COMPLETE LEVEL B											
	1 - 14	15 - 29	30 - 43	44 - 57	58 - 71	72 - 85	86 - 99	100 - 113	114 - 127	128 - 141	142 - 165	166 - 179
Frequencies for Observed Sample S_{60}	3	4	5	2	7	6	8	9	5	5	2	4
Frequencies for Simulated Sample S_{600}	90	60	58	102	61	82	45	65	20	17		
Cumulative Distribution of Observed Sample S_{60} (X)	$\frac{3}{60}$	$\frac{7}{60}$	$\frac{12}{60}$	$\frac{14}{60}$	$\frac{21}{60}$	$\frac{27}{60}$	$\frac{35}{60}$	$\frac{44}{60}$	$\frac{49}{60}$	$\frac{54}{60}$	$\frac{56}{60}$	$\frac{60}{60}$
Cumulative Distribution of Simulated Sample S_{600} (X)	$\frac{90}{600}$	$\frac{150}{600}$	$\frac{208}{600}$	$\frac{310}{600}$	$\frac{371}{600}$	$\frac{453}{600}$	$\frac{498}{600}$	$\frac{563}{600}$	$\frac{583}{600}$	$\frac{600}{600}$	$\frac{600}{600}$	$\frac{600}{600}$
S_{60} (X) - S_{600} (X)	$\frac{60}{600}$	$\frac{80}{600}$	$\frac{88}{600}$	$\frac{170}{600}$	$\frac{161}{600}$	$\frac{183}{600}$	$\frac{148}{600}$	$\frac{125}{600}$	$\frac{93}{600}$	$\frac{60}{600}$	$\frac{40}{600}$	0

$$D = \text{Maximum} \quad S_{60}(X) - S_{600}(X) = \frac{183}{600} = .3500$$

Level of Significance at .05 = 1.36 (.1365) = .1856

Table 4.14

Data Cast for Kolmogorov-Smirnoff Two Sample Test for Compariosn of Observed Sample with Simulated (Predicted) Samples for Level D (1968-69 & 1969-70)

	TIME IN DAYS TO COMPLETE LEVEL D													
	13 - 32	33 - 52	53 - 72	73 - 92	93 - 112	113 - 132	133 - 152	153 - 172	173 - 192	193 - 212	213 - 232	233 - 252	253 - 272	273 - 292
Frequencies of Observed Sample S_{99}	0	0	7	5	9	17	12	9	9	3	5	18	2	3
Frequencies of Simulated Sample S_{990}	52	90	151	210	202	144	74	42	14	11	0	0	0	0
Cumulative Distribution of Observed Samples $S_{99}(X)$	$\frac{0}{99}$	$\frac{0}{99}$	$\frac{7}{99}$	$\frac{12}{99}$	$\frac{21}{99}$	$\frac{38}{99}$	$\frac{50}{99}$	$\frac{59}{99}$	$\frac{68}{99}$	$\frac{71}{99}$	$\frac{76}{99}$	$\frac{94}{99}$	$\frac{96}{99}$	$\frac{99}{99}$
Cumulative Distribution of Simulated Samples $S_{990}(X)$	$\frac{52}{990}$	$\frac{142}{990}$	$\frac{293}{990}$	$\frac{503}{990}$	$\frac{705}{990}$	$\frac{849}{990}$	$\frac{923}{990}$	$\frac{965}{990}$	$\frac{979}{990}$	$\frac{990}{990}$	$\frac{990}{990}$	$\frac{990}{990}$	$\frac{990}{990}$	$\frac{990}{990}$
$S_{99}(X) - S_{990}(X)$	$\frac{52}{990}$	$\frac{142}{990}$	$\frac{223}{990}$	$\frac{383}{990}$	$\frac{495}{990}$	$\frac{469}{990}$	$\frac{423}{990}$	$\frac{375}{990}$	$\frac{299}{990}$	$\frac{280}{990}$	$\frac{230}{990}$	$\frac{50}{990}$	$\frac{30}{990}$	0

$$D = \text{Maximum} \quad S_{99}(X) - S_{990}(X) = \frac{495}{990} = .5000$$

Level of Significance at .05 = 1.36 (.1005) = .1366

this level. However, it could be a change in attitude of the prescribers or an attempt to better reinforce material as it is learned. A user highly involved in the IPI system would be aware of this type of change and should be able to adjust parameters in order to improve the prediction ability of the model. It could well be that a model such as EDSIM 1 would have excellent predictive ability when combined with the insights of a user who is very knowledgable of the system.

Examination of Student Attributes

One of the secondary objectives of this study was to examine possible relationships between student background data and time to completion of a specific unit or level. Over the past few years other studies have been conducted to examine various types of learning rate measures and the relationship between these rate measures and selected student characteristics. Essentially, all of these studies had failed to evidence any relationship between selected student characteristics such as intelligence, unit pretest scores, past achievement in reading and mathematics, age and various measures of the rate of learning under the system of Individually Prescribed Instruction. (Glaser, 1968; Lindvall and Bolvin, 1967).

This phase of this study was concerned with characteristics that would be known about a student on his initial entry into a level. Background data was available on age, grade, intelligence quotient, mathematics placement scores, unit pretest scores, total units previously mastered, and standardized achievement scores in mathematics. Samples of students that had participated in specific levels were selected. The background data that was available for these students were chosen as the independent variables to be correlated with the dependent variable of number of days needed to complete a unit or level. Most of the samples were concerned with days to complete a specific unit since samples of students that had completed levels were usually too small to make realistic samples. Correlation matrices were run to identify the most likely background data that could have effect on the dependent variable. Examination of the matrices indicated that only two variables had effects that were constant over units and levels. These two variables were age and unit pretest scores. It was also found that the ID number would correlate highly with time in a unit. However, the majority of ID numbers were assigned so as to reflect the age and grade of the student which would explain the relationship.

Placement tests unlike unit pretest scores had very little constant effect over units and levels. The geometry placement test scores did correlate highly with time in the units in Level D but were not consistent in their relationships over units and levels. In fact, no standardized test scores could be found to have a constant relationship to time in a level or unit. It appeared reasonable to assume that any norm referenced measure would not correlate with time in a unit since each unit had very specific performance criteria that it was designed to measure. Unit pretests were designed to measure specific knowledge in relation to the criteria in a unit and did correlate highly with time needed to complete a unit.

Of all the background data examined, IQ had the least constant effect over units and levels. It would appear to be one of the poorest predictors for determining time in a unit or level.

It was also found that the number of units previously mastered by a student was inconsistent in its correlation over units and levels. It would sometimes have a very positive effect (+.43) in one unit and then have a negative effect (-.39) for that same unit in a different level.

Multiple R's were computed for selected samples in selected units. Significant R's could be found in all units and levels (.80 for D3; .68 for E4) but the factors that contributed to the multiple R were not the same over units or levels.

This examination of relationships was being performed in order to explore the possibility of adding student attributes to the model. This type of addition should then make it possible to have students with certain characteristics sample a smaller section of unit service time distributions. It would also make it possible to assign different pre-test and post-test probabilities to students with identifiable characteristics. However, if factors could not be found that were constant over units and levels, input into the model would be difficult.

It is evident from this research and previous research (Yeager and Kissel, 1969) that the number of days a student requires to master a unit is related to the students initial entering state. Unit pretest scores, age, and skills that must be mastered have been shown to be important factors. Although beta weights associated with variables change in relationship to a particular unit, further study could be made to examine the relative stability of these weights over successive years for a given unit.

It must also be stated that an effective student guidance subroutine for use in a simulator such as EDSIM 1 would probably need to include personality factors. It is quite possible that the aggressive child with average ability may move more rapidly through a unit of IPI than the timid child with above average ability and achievement. This area of research was beyond the scope of this study but could easily be examined if appropriate data were collected.

CHAPTER V

CONCLUSIONS AND RECOMMENDATIONS

This chapter includes the concluding statements concerning the overall descriptive and predictive ability of the EDSIM 1 model. It also includes a section concerned with the feedback that has been received from educators in regard to possible uses of a model such as EDSIM 1. As a result of the experience gained in this study, observations and recommendations are included concerning data management, statistical tests as they might apply to simulation validation, and the overall timing and scope of the validation process in relation to simulation models.

Descriptive Validity

It can be concluded that the variables used in EDSIM 1 are sufficient to describe time to completion of levels in an educational program such as IPI. The model does lack the ability to produce enough extreme results since repeated sampling of service time distributions tends to pull the accumulated service times toward the average.

The descriptive ability of the model is better in higher levels of the program (D and E) than in the lower levels (B and C). Observed samples of students that complete a level indicate that the trend toward a bimodal distribution is greater in the lower levels (B and C). The observed samples from Levels D and E indicate a trend toward a more normal distribution. Since any simulated sample from a model such as EDSIM

1 would have a normal curve as its underlying distribution, it would be natural to expect better descriptive capabilities from the model for Levels D and E. The statistical difference between the observed and simulated samples using the improved model for Level E was very small (.0871) as compared to a rather large difference for Level B (.1538).

Tests of sensitivity indicate that the model is sensitive to small differences in input parameters. The descriptive tests run in this study were based on careful parameter estimation from historical data. A strong relationship exists between parameter estimation and synthetic output of the model; intuitive or unfounded estimates for variable parameters would be reflected in the output of the model and would reduce the confidence one might have in its ability to describe the system.

Parameter estimations, taken from students that participated in a level, produce simulated samples of students that take longer to complete a level within a year than observed samples of "completing" students. However, the major problem of trying to make a judgment of this trend lies in finding realistic "completing" samples of students that complete a level within one year. Some of the students, as stated previously, were probably encountering the material as review or reinforcement material and thus appear to take shorter completion times than is true in the real world. Better and more realistic samples of "completing" students could be found when two years of data were combined for description across years. The statistical difference between simulated and observed distributions is greater for samples across years but these samples are probably more representative of the real world.

Much of the judgment of the validity of a model such as EDSIM 1 must be qualitative rather than quantitative. The problem of finding realistic "observed" samples for use in a comparison with simulated samples could produce less reliable results than is usually the case in other statistical applications. The range of days to completion of a level may, in fact, be one of the most realistic statistics that could be used. However, when statistical tests were applied under the assumption that samples taken from historical data were "realistic" completing samples that could be used for comparison with simulated samples, it was shown that the two samples had distributions of completion times that were not significantly different from each other. This indicates that the model is capable of producing a simulated sample that has a comparable distribution of completion times to the observed sample.

Predictive Validity

Although the model shows ability to describe the system under study, it demonstrates little ability to predict future time distributions from parameter estimates established from past data. This could be considered enough by most writers and builders of models to label the model as invalid. However, a system that operates for the purpose of allowing students to move at individual rates toward mastery of specified performance criteria, is not the type of system that is likely to produce stable variable parameters. Therefore, it is not a truly predictable system. IPI is this type of system.

The model could then be considered as furnishing pertinent information concerning the system operation. It indicates the degree of change that has occurred in time distributions within the system. Arguments could be presented that such information is only descriptive of system behavior and is not predictive. If, in fact, the model has no predictive ability with its present variables, is it a worthwhile model to pursue by further research? This brings up the question of the usability of the model which is discussed in the next section.

Possible Uses of the Model

Simulation as a practical tool of analysis has developed enormously in recent years. If this trend continues, many managers, educational, as well as business and economic, could well begin to occupy their time by developing new computer information systems and decision rules and testing their ideas through simulation. Should this occur, it will require somewhat different skills and training than are presently being required of educational administrators.

EDSIM 1 could well be used as a training model. It is a very inexpensive model to operate. The average time for a run is approximately thirty-five seconds on the CDC 3600 Computer at the University of Massachusetts. The model is also straight-forward and easy to adjust both in output formats and in number of variables that can be considered.

Its major usefulness in this area could be to introduce educators to the input-output relationship of system operation. It could also be used to explore and demonstrate the man-machine interactions that are very fundamental to good simulation procedures. The EDSIM 1 model is an excellent basic model with which educators could conduct experiments that could lead to the development of more complex and possibly more useful models.

The EDSIM 1 model could easily be placed on a time sharing terminal for easy use by educators. This would allow ready access and make it easier for educators to experiment with the model. In many cases, educational administrators approach simulation procedures with distrust, fear, and open hostility. Their next reaction is usually one of awe in view of its ability to produce quick and impressive looking output. It is only after repeated exposure and exploratory use that educators can begin to appreciate its true potential as a decision aiding tool and can begin to appreciate the two-way communication that is possible between man and model.

The model has great potential in helping administrators begin to think through the operational characteristics of new programs being planned. This use is in keeping with the original purpose for which the model was created. In this area of use, the model helps to focus attention on how much (or little) is known about the activities being planned. Since the major variable considered in the model is time, it can help to indicate when the apparent time requirement for students in specific areas is unrealistic.

In dealing with time in a system the model could help educators to identify the variables crucial to the operation of the system. Educational planners are many times more concerned with the psychological and philosophical aspects of the plan than with the operational aspects. Evaluation of the system operation is in most cases done by focusing on the people observing the operation rather than on the actual measurable components of the operation. A model such as EDSIM 1 has great potential in forcing educators to focus on the facts rather than on the people observing the facts. Iteration of the model, with one parameter change on one variable, and every other aspect of the input held constant, can produce an observable interaction in the system operation that is not dependent on perceptive bias.

The EDSIM 1 and related models have been exposed slowly to the academic world, primarily to obtain feedback regarding potential usefulness, and secondarily to find selected situations for additional applications of one or more of the models. Professional meetings include:

Association for Educational Data Systems
Annual Convention, May, 1970: mention of models in session on Simulation Modeling
Annual Convention, April, 1971: major presentation given

American Educational Research Association
Pre-session on Operations Analysis of Education, February, 1971. Equivalent of one full day of presentations and laboratory exercises.

Staff Personnel Utilization Project - Leadership Training Institute, October, 1970, Florissant, Colorado. One-half day equivalent model presentation and elementary laboratory work.

Beaverton, Oregon, Differentiated Staffing Project.
Three day staff seminar, January, 1971.

In addition, two consultants have reacted to the EDSIM 1 and related models. Mr. Joseph B. Crawford of the Westport, Connecticut schools, whose main concern is implementing Programmed Budgeting; and Dr. Fred O. Pinkham, former Director of Project Yardstick in Cleveland, Ohio, and before that, President of Ripon College, have provided their advice.

The following points have come out of these activities and consultations:

1. The average educational administrator is concerned with different problems than those approached by EDSIM 1 and related models. Such questions as probable numbers of students and probable dollar availability and need in future years are more immediately important.
2. The models as they exist are more appropriate for planning massive changes than for stable or slowly changing situations. There is, apparently, little immediate applicability to existing traditional situations. General opinion of the consultant is that the average school man is more numerous than any other kind; he is the ultimate client for this work.
3. If models such as these are to be supported in the academic marketplace, they need to be made more useful in existing situations. The alternative is to support model development and use as a research and development tool, until such time as many more institutions are making massive changes.

4. There may be merit in considering models such as these as training tools. If such models have enough basic, analytic thinking behind them, and have clear, comprehensive instructions for use, it may be possible to start school personnel analyzing their own programs in operational terms. The process of fitting one's own program to various models may help develop a better analytic understanding of that program, regardless of the models used or the results obtained from using them. In essence, this is a thought process more than a modeling process.

Upon completion of the research done in this study, the results were presented to the staff of the Learning Research and Development Center in Pittsburgh, Pennsylvania. This center is responsible for the initial design of the IPI system and for its continuing development.

The following points were brought out in the discussion:

1. It was the general agreement of the group that no major misrepresentation had been made in extracting the IPI data for application to the EDSIM 1 model.
2. It was the opinion of the group that the EDSIM 1 model in its present form and without further improvements would be of little practical use to the IPI project as it now exists. It was suggested that operation of the model from the detail of units and skills rather than from levels and units could enhance its usefulness to IPI.

3. There was agreement that the findings concerning student attributes found in this study were in agreement with those findings made by the staff. The only attribute that seemed to have a consistent effect on rate of learning was age. It was suggested, however, that research should be done to study the effect of age over successive units and these findings could be incorporated into the operation of the model.
4. It was also suggested that the most likely factor that could be used as a predictor would be classroom management procedures. There was a strong feeling that the overall attitude of the teacher and the student-teacher interaction factor greatly affected rate of learning. This was suggested as an area for further research.
5. It was suggested that Research for Better Schools, Incorporated, the regional research laboratory in Philadelphia, Pennsylvania, could probably suggest IPI pilot schools that could be used for further research with the EDSIM 1 model. It was the general agreement of the group that IQ and units previously mastered could well prove to be predictors in the model if examined from these schools. It was also felt that the model might predict time to completion more accurately in these schools, since they were not as developmental in nature as the Oakleaf IPI system.
6. The group indicated that research had been done to see if personality of the student affected rate of learning. It was

generally agreed, although not unanimously, that the personality variables had not proven to be predictors of learning rate. This examination of personality variables had been suggested earlier in this study as a recommendation for further research.

7. The fact that the simulated distribution had not matched very closely the observed distribution in the prediction phase of the model was accepted as a realistic description of the system operation. It was explained that certain changes in regard to more standard procedures in assigning prescriptions according to scores on unit pretests had been initiated in 1969-1970. This could have accounted for the difference in the observed and simulated distributions.

8. The continuing development of IPI has led to the addition of the Primary Education Program (PEP) that is now used as an extension to the lower levels of IPI. The total program is now known as Instruction, Design, and Evaluation (IDE). There was interest expressed by Dr. Resnick, Co-director of IDE, for the development of a simulation that could be used in developing operational techniques for the individual testing program that is a part of PEP.

Statistical Tests in Model Validation

Choice and application of statistical tests for comparison of simulated and observed data is an area that needs extensive research. At present, there are almost no precedents to follow. If simulation is to become a much used process, better statistical procedures will need to be developed for judging the model's accuracy.

This study indicates that graphing techniques may be one of the most effective statistical procedures that can be used. These techniques allow the user to get an overall view of what is happening in the system description.

Recommendations can be found in the current literature for using the Chi Square Goodness of Fit Test for the comparison of simulated and observed data. Experiments with this statistic in this study demonstrated that it was totally ineffective. Observed samples were usually small and thus expected frequencies were small for use with the Chi Square Test.

This means that cell frequencies must be collapsed so as not to violate the application of the test. In most cases, this involved collapsing a distribution of twelve time intervals into one of six or four. The distribution was then incapable of furnishing any information concerning agreement of distributions for simulated and observed samples.

The fact that the underlying distribution of any of the simulated samples in this model could be considered normal was a very difficult problem to handle statistically. It was observed, however, that the best applications available for comparison of distributions were to be found among non-parametric statistics.

Data Management

Data management for the validation study is a very real problem. In order to properly keep track of results from the runs of the model punch statements had to be added to the model so that data could be taken out on data cards. The printed material produced in such a study is overwhelming.

A filing and expectation system is necessary. That is, one needs to decide what he expects to get from certain runs. This one aspect may be the only information retained from this run. This must be marked and filed for easy future reference. If the entire run is kept, a storage room may be necessary.

It is also wise to purge the data occasionally. Many time inputs that are very comparable have been prepared. The most appropriate one must be marked and filed or much time is spent looking through boxes of cards that all appear to be the same thing and contain the exact same information.

Since improvement may be made to the model, it is wise to mark one deck as the original source deck. This deck is used for reference and is never changed. Improvements that are made are all included in another deck. This way, you are better able to judge the relative effects of the improvements. It is always wise, of course, to have binary decks punched for source models if many runs are to be made. If this is not done, the printed matter collects more quickly because of the lists of the source decks that are produced. However, old binary decks that no longer represent a source model of interest should be discarded. Otherwise, they have a tendency to collect and cause a filing problem.

Validation As a Separate Process

The most vexing questions about any simulation model are, "How do you know it is valid?", and "Will it predict accurately?". These are questions that must be answered at some point in any simulation study. However, from experience gained in this study, it is not recommended that the validation study be conducted separate from the process of the model building. Most of the information garnered in iteration of the model in regard to sensitivity tests, internal programming, and random number generation would be of great concern to the model builder. If this aspect of the model is well documented, no reason exists for repetition of these processes. Corrections of poor assumptions made in defining variables in the model can be made easier during the model building process than during the validation process if data is available. No one program ideas and concepts in the same way and revision of another's program is much more difficult than writing an original program.

Model builders should make an attempt to validate all concepts built into the model with historical data, if possible. When it is neither possible to get historical data nor current data, the model builder should attempt to get feedback on the concept he intends to model from those that have been involved in a comparable system or that might be involved in such a system. If the system involves educators, then educators should be invited to participate. Of course, certain validation procedures will be necessary after the model is built but, it should not be considered only as separate process from model building.

The increased use of experimentation of models by applying them to existing systems should prove to be more valuable to educators in the development of simulation as a planning and analytic tool than attempts at judging a model's validity. In most cases, if the model can be proven to be useful, it has also been proven to be valid. The behavioral aspect of educational institutions may make it difficult for them to use simulations with great effect. While the EDSIM 1 model can model the interaction between certain observable components in a curriculum, it seems doubtful that predictors will be found to model the behavior of students interacting with the curriculum.

Simulation in business and economics has permitted the explorations of inventory systems, job shop scheduling, etc. It makes it possible to explore the consequences of the interactions of a large number of processes. These simulations can then predict such system variables as stock shortages, operating costs, or delivery problems. Data for these models can be easily obtained by observation of the many subsystems. However, when students react with teachers, teachers with administrators, and administrators with ego, etc. it is difficult to observe all the subsystems. Many variables that need to be known are inside the human system and cannot be observed.

These problems of simulation occur not only in the study of a single individual but in studies of groups of individuals. Eventually, we may be able to understand the communication between individuals, since this aspect is observable in most cases. However, that part of the process that goes on in the individual's head is going to be a difficult aspect to observe. This problem is recognized by many writers in the field of simulation, and it is therefore generally conceded that models that deal with behavior have yet to demonstrate any concrete validity. "Insofar as their validity is concerned, it is premature to reject or accept the value of most simulations and games in the behavioral sciences." (Hermann, 1967)

BIBLIOGRAPHY

Anderson, G. E., Jr. Adult Resources Flow Model. Unpublished simulation model, School of Education, University of Massachusetts.

Beaird, J. H. and Standish, J. T. Audio Simulation in Counselor Training. U.S. Office of Education, NDEA Title VII Project Number 1245. Monmouth, Oregon: Teaching Research Division, Oregon State System of Higher Education, 1964.

Bonini, Charles P. Simulation of Information and Decision Systems in the Firm. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1963, Chapter 7.

Bonini, C.; Jaedicke, R. and Wagner, H. (eds.) Management Controls: New Directions in Bosie Research. New York: McGraw-Hill, 1963.

Brown, R. Smoothing, Forecasting, and Prediction of Discrete Time Series. Englewood Cliffs, New Jersey: Prentice-Hall, 1963.

Bulkin, M. W. et al. "Load Forecasting, Priority Sequencing, and Simulation in a Job Shop Control System," Management Science, Volume 12, Number 7, 1966.

Churchman, C. West. "An Analysis of the Concept of Simulations." Symposium on Simulation Models. Austin C. Haggatt and Frederick E. Balderston (editors). Cincinnati: Southwestern Publishing Company, 1963.

Churchman, C. "Managerial Acceptance of Scientific Recommendations." California Management Review, Volume 7, Number 1, Fall, 1964.

Cogswell, John F. "Systems Analysis and Computer Simulation in the Implementation of Media." Audio-Visual Instructor, 1965, Volume 10, pp. 384-386.

Conway, R. W. "Some Tactical Problems in Digital Simulation." Management Science, Volume 10, Number 1, October, 1963.

Conway, R. W., Johnson, B. M. and Maxwell, W. L. "Some Problems of Digital Systems Simulation." Management Science, Volume 6, Number 1, October, 1959.

Cox, D. and Lewis, P. The Statistical Analysis of Series of Events. New York: John Wiley and Sons, 1966.

Cruickshank, Donald E. "Simulation: New Directions in Teacher Preparation." Phi Delta Kappan, Volume 48, 1966, pp. 23-24.

Emshoff, James R. and Sisson, Roger L. Computer Simulation Models. New York: The Macmillan Company, 1970.

Frijda, N. "Problems of Computer Simulations." Behavioral Science, Volume 12, January, 1967.

Geisler, Murray A. "Appraisal of Laboratory Simulation Experiences." Management Science, Volume 8, Number 3, April, 1969.

Greenberger, M. (ed.) Computers and the World of the Future. Cambridge, Massachusetts: M.I.T. Press, 1962.

Greenlaw, P. S.; Herron, L. W. and Rawdon, R. H. Business Simulation in Industrial and University Education. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1962.

Guetzkow, Harold (ed.) Simulation in Social Science: Readings. Englewood Cliffs, New Jersey: Prentice-Hall, 1962.

Guetzkow, Harold, et al. Simulation in International Relations: Development for Research and Teaching. Englewood Cliffs, New Jersey: Prentice-Hall, 1963.

Hermann, C. "Validation Problems in Games and Simulations." Behavioral Science, Volume 12, May, 1967.

IBM. Bibliography on Simulation. Report 320-0924-0. IBM, White Plains, New York, 1966.

Jennings, Norman H. and Dickins, Justin H. "Computer Simulation of Peak Hour Operations in a Bus Terminal." Management Science, Volume 5, Number 1, October, 1958.

Kaczka, Eugene E. and Kirk, Roy V. "Managerial Climate, Work Groups, and Organizational Performance." Administration Science Quarterly, Volume 12, Number 2, September, 1967.

Kidera, Edward H. and Hoff, Jean M. "Simulation: Management Tool in Decision-Making." Automation, February, 1968.

King, Paul E. "Simulation: Management Planning Tool." IEEE Transactions on Systems Science and Cybernetics, Volume SSC-4, Number 4, November, 1968.

Kuehn, A. A. and Hamburger, M. J. "A Heuristic Program for Locating Warehouses." Management Science, Volume 9, Number 3, September, 1963.

Lindvall, C. M. and Cox, Richard C. Evaluation as a Tool in Curriculum Development: The IPI Evaluation Program. American Educational Research Association Monograph Series on Curriculum Evaluation, Number 5. Chicago: Rand McNally and Company, 1970.

Lipson, J. I.; Cohen, Henry B. and Glaser, Robert. "The Development of an Elementary School Mathematics Curriculum for Individualized Instruction." Working Paper Number 7, Learning Research and Development Center. University of Pittsburgh, 1966.

Malcolm, D. A. "Bibliography on the Use of Simulation in Management Analysis." Operations Research, Volume 8, March, 1960. pp. 169-177.

McCormich, E. J. Human Factors Engineering. (2nd edition). New York: McGraw-Hill, 1964.

Meier, Robert C. et al. Simulation in Business and Economics. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1969.

Morgenthaler, George W. "The Theory and Application of Simulation in Operations Research" in Progress in Operations Research, Volume 1, Russell L. Ackoff, ed. New York: John Wiley and Sons, Inc., 1961.

Naylor, Thomas H.; Balintfy, J. L.; Burdick, D. S. and Kong Chu. Computer Simulation Techniques. New York: John Wiley and Sons, Inc., 1966.

Naylor, Thomas H. and Burdick, Donald S. "Design of Computer Simulation Experiments for Industrial Systems." Communications of the ACM, Volume 9, Number 5, May, 1966.

Naylor, Thomas H.; Burdick, Donald S. and Sasser, Earl W. "Computer Simulation Experiments for Economic Systems: The Problem of Experimental Design." American Statistical Association Journal, Volume 62, Number 320, May, 1966.

Naylor, Thomas H. and Finger, J. M. "Verification of Computer Simulation Models." Management Science, Volume 14, Number 2, October, 1967.

Richards, Thomas C. An Analysis of Resource Allocation in Planning a Performance Criteria Curriculum. Unpublished Ph.D. dissertation. University of Massachusetts. December, 1970.

Rowe, Allan J. "Computer Simulation: A Technique for Management Problems." Proceedings - Fall Joint Computer Conference. New York: Spartan Books, 1968.

Siegel, Sidney. Non-parametric Statistics for the Behavioral Sciences. New York: McGraw-Hill Book Company, 1956.

Smith, R. and Greenlaw, P. Simulation of a Psychological Decision Process in Personnel Selection." Management Science, Volume 13, Number 8, April, 1966.

Steer, D. J. and Page, A. C. "Feasibility and Financial Studies of a Part Installation. Operational Research Quarterly, Volume 12, Number 3, September, 1961.

Teichroew, Daniel. "History of Distribution Sampling Prior to the Era of the Computer and Its Relevance to Simulation." American Statistical Association Journal, Volume 60, Number 309, March, 1965.

Twelker, Paul A. "Simulation: An Overview." Monmouth, Oregon: Teaching Research Division, Oregon State System of Higher Education. ERIC 025 459, 1969. (Mimeographed.)

Yeager, John L. and Kissel, Mary Ann. "An Investigation of the Relationship Between Selected Student Characteristics and Time Required to Achieve Unit Mastery." Working Paper 46. Learning Research and Development Center. University of Pittsburgh, 1969.

Zelditch, Morris, Jr. and Evan, W. H. "Simulated Bureaucracies: A Methodological Analysis" in H. Guetzkow (ed.) Simulation in the Social Sciences: Readings. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1965.

APPENDIX

The Computer Program

The Computer Source Program that was used in this study follows.

The original computer program was created primarily by George F. Williams and Wayne E. Leininger, under the guidance of Dr. Eugene E. Kaczka, all of the School of Business Administration, University of Massachusetts.

The program that is included in this appendix is the modified version of the original with modification made by William W. Foley and Frederick A. deFriesse of the School of Education, University of Massachusetts.

DICTIONARY

ALOW: Lowest time in run

APOST: Probability of passing a post-test

APRE: Probability of passing a pretest

BLOCK: Used to set up numbers in each interval of the histogram

DCUM: Cumulative distribution for unit service time distribution before a post-test

DEM: Demand for each event type by student

ETDIST: Event type distribution or unit service time distribution in discrete probabilities for before a post-test

ETDIST2: Event type distribution or unit service time distribution in discrete probabilities for after a post-test

ETSUM: Total demand for each event type in each area

ETIME: Time in hours or days for each event type.

FAIL: Number of students that fail a post-test

HIGH: Highest time in run

IPRO: Used as counter for number of PC's in each area

ISUB: Counter for intervals in histogram

KPD: Counter for those that pass pretest

KFD: Counter for those that fail post-test

NOET: Number of event types in a service time distribution before a post-test

NO2ET: Number of event types in a service time distribution after a post-test

PASS: Number that pass a pretest

PCINA: Number of performance criteria that are in each are at the beginning of the run

PERFAIL: Percent that fail a post-test

PERPAS: Percent that pass a pretest

RNOG: Random number function

STIME: Totatl time for a student in a run

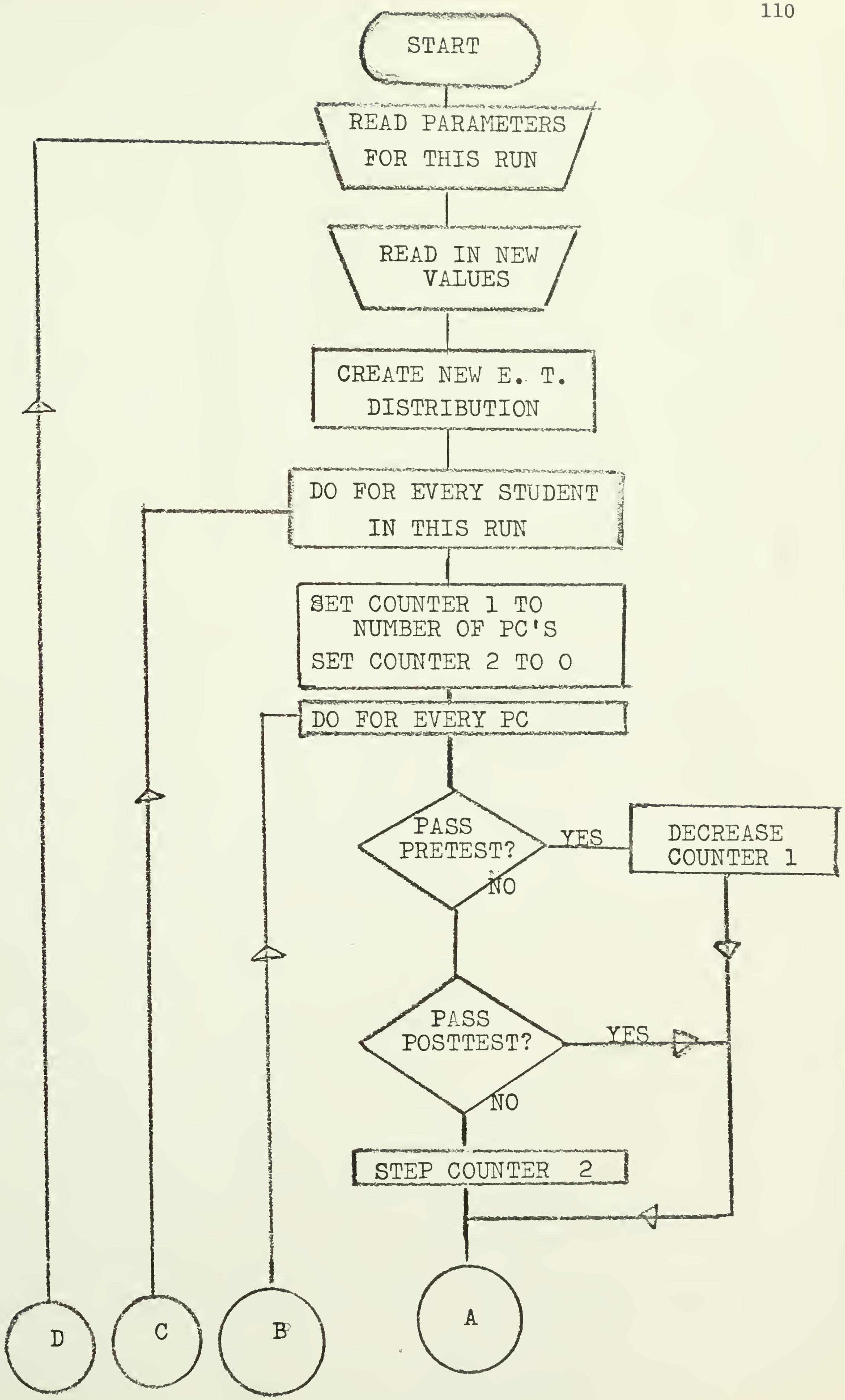
STUD: Number of students in a run.

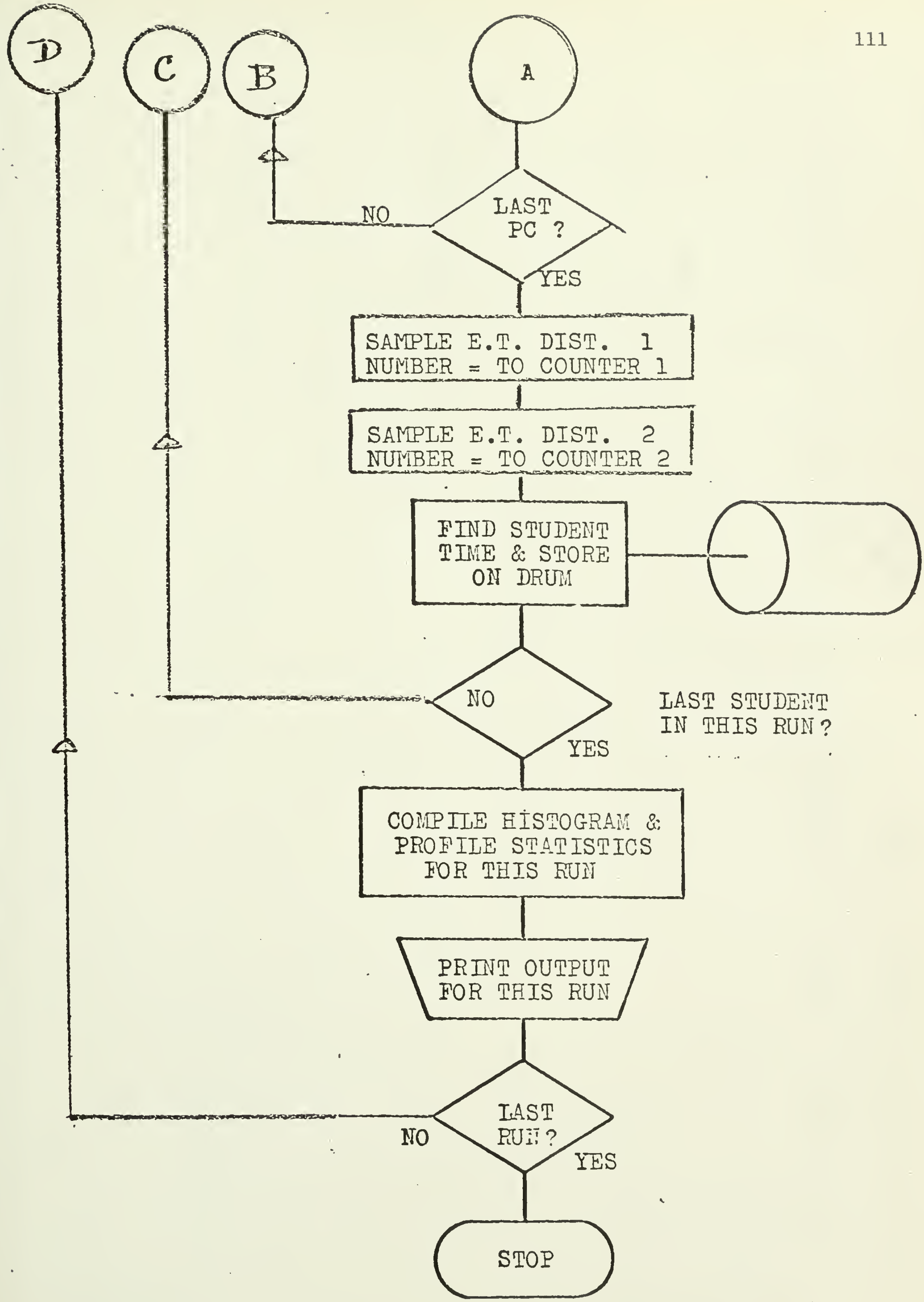
SUB: Intervals for histogram

TMEAN: Mean service time for a unit (area)

TPCINA: Total number of performance criteria required of each student after pretesting and posttesting

TDIF: Range of time in a run


START
READ PARAMETERS FOR THIS RUN
READ IN NEW VALUES
CREATE NEW E. T. DISTRIBUTION
DO FOR EVERY STUDENT IN THIS RUN
SET COUNTER 1 TO NUMBER OF PC'S
SET COUNTER 2 TO 0
DO FOR EVERY PC
PASS PRETEST?
YES
DECREASE COUNTER 1
NO
PASS POSTTEST?
YES
NO
STEP COUNTER 2
D
C
B
A


D
C
B
A
LAST PC ?
NO
YES
SAMPLE E.T. DIST. 1
NUMBER = TO COUNTER 1
SAMPLE E.T. DIST. 2
NUMBER = TO COUNTER 2
FIND STUDENT TIME & STORE ON DRUM
NO
YES
LAST STUDENT IN THIS RUN?
COMPILE HISTOGRAM & PROFILE STATISTICS FOR THIS RUN
PRINT OUTPUT FOR THIS RUN
LAST RUN?
NO
YES
STOP

```
      PROGRAM FDSIM1
      COMMON /A/ PCINA(12,13), APRE(13), APOST(13), TPCINA (13),
     1PASS(13,10), FAIL(13,10), DCUM(13,25), ETSUM(13,25), NOET,
     1SCUM(13,25),NO2ET
      DIMENSION FTIME(25), ETDIST(13,25),TSUM(13),TMEAN(13)
      DIMENSION BLOCK(10),COUNT(10),FROM(10),TO(10)
      DIMENSION DEM(13,25), TDEM(25), WORD(10)
      DIMENSION PERPAS(13,10), PERFAIL(13,10)
      DIMENSION      PTLE(20,5),ETDIST2(13,25)
      DIMENSION  KFD(13),KPD(13)
      INTEGER DEF, PROFILE,PEECEE,AREAS
      INTEGER PROLIM
      FAKE = RNOG(1)
      RUN = 0.0
C     READ PARAMETERS FOR THIS RUN
    3 READ 201,ISTUD,NOET,NO2ET,PROLIM,DEF
      IF(EOF,60) 800,4
C     READ IN NEW VALUES
    4 CONTINUE
      CALL DEFINE(DEF,PROFILE,PEECEE,AREAS)
      RUN = RUN + 1.    $    KP = 0
      STUD = ISTUD
      READ 209, (WORD(J),J=1,5)
      DO 5 M=1,PROLIM
    5 READ 210,  (    PTLE(M,I),I=1,3)
      READ 203, ((PCINA(I,J),J=1,13),I=1,PROLIM)
      READ 203, (APRE(J),J=1,13)
      READ 203, (APOST(J),J=1,13)
      READ 207, (ETIME(J),J=1,25)
      READ 205, ((ETDIST(I,J),J=1,25),I=1,13)
      READ 205, ((ETDIST2(I,J),J=1,25),I=1,13)
C     HEADING OUTPUT
      KP = KP + 1   $    PRINT 220,KP
      PRINT 302
      PRINT 304, RUN
      PRINT 306, (WORD(J),J=1,5)
      PRINT 308,PROFILE,ISTUD
      PRINT 310,PROFILE,PROLIM
      PRINT 312
      PRINT 230
      PRINT 314, (APRE(J),J=1,13)
      PRINT 315, (APOST(J),J=1,13)
C     CREATE NEW CUMULATIVE E.T. DIST
      DO 17 J=1,13
      DCUM(J,1) = ETDIST(J,1)
      SCUM(J,1)=ETDIST2(J,1)
      DO 15 I=2,NOET
   15 DCUM(J,I) = DCUM(J,I-1) + ETDIST(J,I)
      DO 16 I=2,NO2ET
   16 SCUM(J,I) = SCUM(J,I-1) + ETDIST2(J,I)
C     DO 17 I=1,NOET
C     SCUM(J,I) = SCUM(J,I) -.001
C  17 DCUM(J,I) =DCUM(J,I)-.001
   17 CONTINUE
C     PRINT 267, ((DCUM(I,J),I=1,13),J=1,NOET)
C     PRINT 271
      DO 400 IPRO=1,PROLIM
C     CLEAR AREA COUNTERS
      DO 11 I=1,13   $    TSUM(I)=0.
      DO 11 J=1,10   $    PASS(I,J)=0.    $    FAIL(I,J)=0.
```

```
   11 CONTINUE
      DO 12 J=1,13
      DO 12 I=1,25    $    TDEM(I) = 0.
   12 DEM(J,I) = 0.
C     LOOP FOR NO. OF STUDENTS
      DO 100 IS=1,ISTUD
C     FOR EACH STUDENT, MARCH THROUGH THE MAZE
C     SUBROUTINE PTEST NOW HANDLES THE AREAS
   25 CALL PTEST(IPRO,KFD,KPD)
C     HAVING COMPLETED PRE/POST, SAMPLE ETDIST
      CALL SAMPLE (KFD,KPD)
      STIME=0.
C     FIND STUDENT TIME AND STORE ON DRUM VIA LU 11
      DO 29 I=1,13  $  CONST=0.  $  DO 27 J=1,NOET
      DEM(I,J) = DEM(I,J) + ETSUM(I,J)
   27 CONST = CONST + ETSUM(I,J) * ETIME(J)
      TSUM(I) = TSUM(I) + CONST
   29 STIME = STIME + CONST
      WRITE(11,251) IS, STIME
      PRINT 253, IS,STIME
C     END STUDENT LOOP
  100 CONTINUE
C     PROFILE STATISTICS SECTION
      REWIND 11
C     SEARCH LIST TWO TIMES FOR HIGH AND LOW
      READ (11,251) IDUM,HIGH
      ALOW = HIGH
      DO 41 I=2,ISTUD
      READ (11,251) IDUM,TEST
      IF(TEST.GT.HIGH) 43,45
   43 HIGH = TEST
      GO TO 41
   45 IF(TEST .LT. ALOW) 47,41
   47 ALOW = TEST
   41 CONTINUE
      REWIND 11
C     SET UP INTERVALS FOR HISTOGRAM
      SUB = 10.
      ISUB = SUB
      TDIF = HIGH - ALOW
      A = TDIF/SUB + .9
      INC = A   $   AINC = INC
      BLOCK(1) = ALOW + AINC - 1.
      DO 61 J=2,ISUB
   61 BLOCK(J) = BLOCK(J-1) + AINC
C     PRINT 301, (BLOCK(I),I=1,10)
      FROM(1) = ALOW
      TO(1) = ALOW + AINC
      DO 65 J=2,ISUB
      FROM(J) = FROM(J-1) + AINC+1.
   65 TO(J) = TO(J-1) + AINC+1.
      DO 48 K=1,10
   48 COUNT(K) = 0.
      T = 0.   $   TSQ = 0.
C     READ IN TIMES FOR HISTOGRAM AND MEAN-SD COMPUTATION
      DO 51 I=1,ISTUD
      READ (11,251) IDUM,TIME
C     COUNT THE TIME IN A HIST INTERVAL
      DO 62 J=1,ISUB
      IF(TIME.LE.BLOCK(J)) 63,62
```

```
62 CONTINUE
63 COUNT(J) = COUNT(J) + 1.
   T = T + TIME
51 TSQ = TSQ + TIME**2
   AMEAN = T/STUD
   DENOM = STUD * (STUD-1.)
   VAR = (STUD*TSQ - T**2)/DENOM
   SD = SQRT(VAR)
   DO 53 J=1,13
53 TMEAN(J)= TSUM(J)/(STUD-PASS(J,1))
   DO 56 J=1,NOET
   DO 56 L=1,13
56 TDEM(J) = TDEM(J) + DEM(L,J)
C     COMPUTE PASS / FAIL OCCURANCES FOR 1 TO 10 TIMES
   DO 67 I=1,13
   DO 67 J=1,10
   PERPAS(I,J) = PASS(I,J) / STUD
67 PERFAIL(I,J) = FAIL(I,J) / STUD
   REWIND 11
C     * * * * * * OUTPUT SECTION * *
C     PROFILE PAGES OUTPUT
   KP=KP+1   $   PRINT 220,KP
   PRINT 345, PROFILE,  IPRO
   PRINT 335,  (    PTLE(IPRO,I),I=1,3)
   PRINT 346,PEECEE,PROFILE
   PRINT 230
   PRINT 348,PEECEE,  (PCINA(IPRO,J),J=1,13)
   PRINT 350
   PRINT 230
   DO 71 K=1,NOET
71 PRINT 352, K, (ETDIST(J,K),J=1,13)
   PRINT 351
   PRINT 230
   DO 70 K=1,NO2ET
70 PRINT 352, K, (ETDIST2(J,K),J=1,13)
   PRINT 354,  PEECEE
   PRINT 355
   PRINT 230
   DO 72 K=1,10
72 PRINT 356, K, (PERPAS(J,K),J=1,13)
   PRINT 358, PEECEE
   PRINT 359
   PRINT 230
   DO 73 K=1,10
73 PRINT 356, K, (PERFAIL(J,K),J=1,13)
   KP=KP+1   $   PRINT 220,KP
   PRINT 375, PROFILE
   PRINT 377, PROFILE,  HIGH
   PRINT 378, PROFILE,  ALOW
   PRINT 380, PROFILE,  AMEAN, SD
C     HISTOGRAM OUTPUT
   PRINT 382, PROFILE, IPRO,INC
   PRINT 384, (FROM(J),J=1,ISUB)
   PRINT 385, (TO(J),J=1,ISUB)
   PRINT 386
   PRINT 388, (COUNT(J),J=1,ISUB)
   PRINT 386
   PRINT 390,  APEAS
   PRINT 230
   PRINT 392, (TMEAN(J),J=1,13)
```

```
382 FORMAT (    47X,*HISTOGRAM FOR*,A5,I5,/,    3X,*WITH INCREMENTS OF*,
   1I4,* HOURS*,//)
384 FORMAT (5X,* FROM *,3X,10(F5.0,4X))
385 FORMAT (/,5X,* TO *,4X,10(F5.0,4X))
386 FORMAT (/5X,8H* -- * -,10(9H * - - * ))
388 FORMAT (/,5X,*NUMBER*,5X,10(F3.0,6X))
390 FORMAT (///,37X,*THE AVERAGE TIME IN HOURS SPENT IN EACH *,A5,/)
392 FORMAT (5X,*HOURS*,3X,13F8.1)
400 FORMAT (45X,*DEMAND FOR EVENT TYPES-WITH*,I3,* TYPES AVAILABLE*,/)
402 FORMAT (5X,*ET*,I2,3X,13F8.0)
404 FORMAT (//,20X,*SUM OF ABOVE DEMAND - BY TYPE*,/)
406 FORMAT (25X,10F8.0)
    END
    SUBROUTINE PTEST(IPRO,KFD,KPD)
    COMMON /A/ PCINA(12,13), APRE(13), APOST(13), TPCINA (13),
   1PASS(13,10), FAIL(13,10), DCUM(13,25), ETSUM(13,25), NOET,
   1SCUM(13,25),NO2ET
    DIMENSION KFD(13),KPD(13)
C     SHIFT PCINA TO TPCINA
    DO 13 J=1,13
 13 TPCINA(J) = PCINA(IPRO,J)
    DO 100 IA=1,13
    P = 0.
    K = 0
    NOPC = PCINA(IPRO,IA)
C     PRETEST SECTION
 10 DO 20 I=1,NOPC
    R = RNOG(2)
    IF(APRE(IA)-R) 20,18,18
 18 TPCINA(IA) = TPCINA(IA) - 1.
    K = K + 1
    IF(K.GT.10) K=10
    PASS(IA,K) = PASS(IA,K) + 1.
    P = P + 1.
 20 CONTINUE
    KPD(IA)=TPCINA(IA)
C     POSTTEST SECTIOM
    F = 0.
    KF = 0
    NOPC = TPCINA(IA)
    DO 30 I=1,NOPC
    R = RNOG(2)
    IF(APOST(IA) - R) 25,30,30
 25 TPCINA(IA) = TPCINA(IA) + 1.
    KF = KF + 1
    IF(KF.GT.10) KF=10
    FAIL(IA,KF) = FAIL(IA,KF) + 1.
    F = F + 1.
 30 CONTINUE
C     PRINT 50, IA, P, F
    KFD(IA)=KF
100 CONTINUE
 50 FORMAT (10X,*FROM SUB TEST  AREA IS *,I5,5X,*PASSING*, F5.0,5X,
   1*FAILURES*,F5.0)
    RETURN  $  END
    SUBROUTINE DEFINE(DEF,PROFILE,PEECEE,AREAS)
    INTEGER DEF, PROFILE,PEECEE,AREAS
    PROFILF=5H
    PFECEF=5H
    AREAS=5H
```

```
      KP=KP+1    $   PRINT 220,KP
      PPINT 400, NOET
      PRINT 230
      DO 74 K=1,NOET
   74 PRINT 402, K, (DEM(J,K),J=1,13)
      PRINT 404
      PRINT 406, (TDEM(J),J=1,NOET)
C     * * * * * END OUTPUT
C     GET A NEW PROFILE
  400 CONTINUE
C     GET A NEW RUN OTHERWISE STOP
      GO TO 3
  800 STOP
C     * * * * * * OUTPUT FORMATS * *
C     * * * * GENERAL FORMATS * *
  201 FORMAT (4I4,2X,A1)
  203 FORMAT (13F5.0)
  205 FORMAT (25F3.0)
  207 FORMAT (5F5.0)
  209 FORMAT (5A8)
  210 FORMAT (2A8,A4)
  231 FORMAT (45X,5A8)
  251 FORMAT (I3,F5.0)
  215 FORMAT (110X,*PAGE NO.*,I3,/)
  220 FORMAT (1H1,110X,*PAGE NO.*,I3,/)
  230 FORMAT (16X,* NUM.    P. V.     ADD.     SUB.    MULT.     DIV.    C.O.P
     2.  FRACT.  MONEY   TIME    MEAS.    GEO.  SP.TOP.*,/)
C     * * * * HEADING PAGE * * * *
  253 FORMAT (2X,I5,3X,F10.4)
  302 FORMAT(58X,*E D S I M  1*,//,36X,*S C H O O L  O F  E D U C A T I
     1O N  S I M U L A T I O N*,6(/))
  304 FORMAT (52X,*RUN NUMBER *,F2.0,///)
  306 FORMAT (30X,*DATA DERIVED FROM *,5A8,//)
  308 FORMAT (20X,*THE NUMBER OF STUDENTS IN EACH *,A5,* IS*,I5,/)
  310 FORMAT (20X,*THE NUMBER OF *,A5, *S FOR WHICH RESULTS WILL BE OUTP
     1UTED IS *,I3,//)
  312 FORMAT (42X,*PRE AND POST TEST PROBABILITIES FOR THIS RUN*,//)
  314 FORMAT (7X,*PRE*,5X,13(2X,F5.3,1X))
  315 FORMAT (/,6X,*POST*,5X,13(2X,F5.3,1X))
  335 FORMAT (57X,2A8,A4,//)
C     * * * * PROFILE PAGES * * * * *
  345 FORMAT (50X,*OUTPUT STATISTICS FOR *,A5,I4,/)
  346 FORMAT (47X,*NUMBER OF *,A5,*S IN THIS *,A5,/)
  348 FORMAT (4X,A5,  3X,13(2X,F5.0,1X))
  350 FORMAT (1H1,/,45X,*EVENT TYPE SELECTION PROBABILITIES*,/,51X,
     1 *BEFORE POST TEST*,/)
  351 FORMAT (1H1,/,45X,*EVENT TYPE SELECTION PROBABILITIES*,/,51X,
     1*AFTER POST TEST*,/)
  352 FORMAT (5X,*ET*,I3,2X,13F8.3)
  354 FORMAT (//,42X,*PERCENTAGES OF STUDENTS PASSING *,A5,* PRETESTS*)
  355 FORMAT (4X,*PERCENT*,/,4X,*PASSING*)
  356 FORMAT (2X,*AT LEAST *,I2,1X,13(3X,F5.3))
  358 FORMAT (//,42X,*PERCENTAGES OF STUDENTS FAILING *,A5,* POST TESTS*
     9  )
  359 FORMAT (4X,*PERCENT*,/,4X,*FAILING*)
  375 FORMAT (43X,*TIME DIMENSION OF THIS *,A5,//)
  377 FORMAT (20X,*HIGHEST TIME IN THE *,A5, * IS*,F8.1, * HOURS*,/)
  378 FORMAT (20X,*LOWEST TIME IN THE *,A5, * IS*,F8.1,* HOURS*,/)
  380 FORMAT (20X,*THE AVERAGE *,A5, * TIME IS *,F8.2,* HOURS WITH A STA
     1NDARD DEVIATION OF *,F7.2,* HOURS*,///)
```

```
      IF (DEF.EQ.1HL)  10, 20
   10 PROFILE=5HLEVEL
      PEECEE=5H AREA
      AREAS=5H AREA
      GO TO 30
   20 IF (DEF.EQ.1HA)  25, 30
   25 PROFILE=5H AREA
      PEECEE=5HSKILL
      AREAS=5HSKILL
   30 RETURN
      END
      SUBROUTINE SAMPLE(KFD,KPD)
      COMMON /A/ PCINA(12,13), APRE(13), APOST(13), TPCINA (13),
     1PASS(13,10), FAIL(13,10), DCUM(13,25), ETSUM(13,25), NOET,
     1SCUM(13,25),NO2ET
      DIMENSION KFD(13),KPD(13)
      DO 8 I=1,13  $  DO 8 J=1,25
    8 ETSUM(I,J) = 0.
C     FOR ALL AREAS, WITH NEW TPCINA
      DO 80 J=1,13
      N=TPCINA(J)
C     GENERATE RANDOM NO. AND FIT INTO CUM DIST.
      M=KPD(J)
      DO 50  I=1,M
      R=RNOG(2)
      DO 30  IDIST=1,NOET
      ISAVE=IDIST
      IF(R-DCUM(J,IDIST))25,25,30
   30 CONTINUE
   25 ETSUM(J,ISAVE)=ETSUM(J,ISAVE)+1.
   50 CONTINUE
      MM=KFD(J)
      DO 65  I=1,MM
      R= RNOG(2)
      DO 60  IDIST = 1,NO2ET
      ISAVE=IDIST
      IF (R-SCUM(J,IDIST)) 63,63,60
   60 CONTINUE
   63 ETSUM(J,ISAVE)= ETSUM(J,ISAVE) +1
   65 CONTINUE
   80 CONTINUE
C     PRINT 70, ((ETSUM(I,J),I=1,13),J=1,NOET)
   70 FORMAT (5X,*FROM SUB SAMPLE  ETSUM*,/,(10X,13F5.0))
      RETURN  $  END
      FUNCTION RNOG(ICODE)
C     GENERATES ONE NO. PER PASS OF LENGTH DETERMINDED BY INDEX
      IF (ICODE.EQ.99999999)  GO TO 100
      RNOG=RANF(-1)
  100 CONTINUE
      RETURN  $  END
         SCOPE
@LOAD
```

The input formats necessary for operation of the EDSIM 1 model follow.

The model may be run using the concept of levels, areas, and units or using the concept of a more detailed description of area, unit, and skills. Depending upon the user's desire, either an L (level) or A (area) is placed in the first parameter card (see input format).

If the user operates the model from concept of level, the number of profiles is usually one (1) since each area in the level consists of a unit. However, if the model is run from concept of areas, the number of profiles is the number of units since each unit would have a different number of skills. This means that a Profile Card (see input formats) must be included for the number of profiles indicated in the parameter card.

The input instructions are listed in terms of the IPI mathematics program. If the model is to be run for other than training purposes, other labels are easy to insert and should be inserted so that output may be filed and referenced with ease.

EDSIM 1 Input Formats, As the Data Deck Should be Submitted

Parameter Card

COLUMN	CONTAINS
1-4	Number of students in run ($\leq$ 9999) I Format
5-8	Number of event types ($\leq$ 25) for before post-test distributions I Format
9-12	Some information for event types for after post-test distributions (usually the same as before post-test distributions but could be different)
15-16	Number of Profiles ($\leq$ 13) I Format
19	L or A for level of detail desired

Profile Card

(one per profile indicated on parameter card)

1-5	Numeration (if operated from detail of <u>level</u> would be 1; if operated from detail of A (area) would be number of skills (objectives) in the unit.
6-10	Place Value
11-15	Addition
16-20	Subtraction
21-25	Multiplication
26-30	Division
31-35	Combination of Processes
36-40	Fractions
41-45	Money
46-50	Time

COLUMN	CONTAINS
51-55	Systems of Measurement
6-60	Geometry
61-65	Special Topics

Pretest Probability Card

(one card required)

1-5	Probability of passing a pretest for numeration
6-10 to 61-65	Some information for other 12 areas listed on Profile Card

Post-Test Probability Card

(one card required)

1-5	Probability of passing a post-test for numeration
6-10 to 61-65	Some information for other 12 areas on Profile Card

Event Type Time Cards

(5 cards are required, even if blank)

		on subsequent card:
1-5	Time in hours to complete event type 1	(6, 11, 16, 21)
6-10	Time in hours to complete event type 2	(7, 12, 17, 22)
11-15	Time in hours to complete event type 3	(8, 13, 18, 23)
16-20	Time in hours to complete event type 4	(9, 14, 19, 24)
21-25	Time in hours to complete event type 5	(10, 15, 20, 25)

All read as F 5.0, but decimal point may be included to give fractional hours.

Event Type Distribution Cards

(13 cards required, 1 per area, in order)

COLUMN	CONTAINS
1-3	Selection Probability, event type 1
4-6	Selection Probability, event type 2
73-75	Selection Probability, event type 3 on through Selection Probability, event type 25
	All read as F 3.0; must be keyed as .xx

More than one set of data may be submitted at a time. The program will operate until an end of file condition is encountered.

The following is an example of the punched card input necessary for a run of the EDSIM I model.

```
 111   25   25    1  L
EDSIM 1 LEVEL D ACROSS YEARS
EDSIM LEVEL D XYEARS CHECK
    1     1     1     1     1     1     1     1     1     1     1     1
  .15   .33   .37   .40   .38   .38   .31   .26   .32   .30   .71
  .44   .37   .50   .51   .61   .40   .31   .44   .42   .38   .63
    1     2     3     4     5
    6     7     8     9    10
   11    12    13    14    15
   16    17    18    19    20
   21    24    29    35    40
.04.05.02.03.03.01.01.01.02.06.04.02.08.08.08.06.03.03.01.02.02.06.10.03.
.05.09.02.03.07.03.00.03.02.03.00.06.06.03.07.04.00.02.08.00.00.10.02.02.
.11.05.06.04.04.05.03.05.05.04.08.05.00.05.05.09.05.00.00.05.03.02
.08.08.04.02.03.02.02.01.06.10.08.00.08.00.04.03.04.06.00.06.09.02.01.01.
.11.10.10.06.03.02.07.09.07.00.06.00.10.00.02.05.02.06.02.01.01
.08.08.08.06.04.04.06.08.04.03.10.00.06.02.02.05.01.02.00.02.01.10
.16.08.09.04.04.06.00.08.00.00.06.02.06.00.08                  .03.03.01.02.02.
.03.03.04.04.04.08.09.10.08.10.10.06.10.07.06.02.00.02.00.02.00.02
.13.13.03.08.08.08.00.04.10.00.11.06.00.02.02.00.04.04.02.02
.15.04.06.08.13.15.09.08.06.06.02.00.02.00.00.02.00.00.00.02.02
.04.08.08.03.00.04.11.00.04.10.00.03.10.00.03.08.05.00.03           .05.06.03.
.61.08.12.03.02.06.08

.25.10.14.10.00.03.00.08.00.10           .08.03           .03.02.03.02
.20.18.10.06.10.05.10.03.00.03.03.06.00.00.02                         .02    .02
.18.22.11.04.02.06.11.02.08.06.00.02.08
.14.34.11.08.10.00.00.06.00.06.00.03.03.00.05
.44.23.14.05.06.08
.19.20.08.08.08.08.08.10   .02   .03         .03          .03
.10.06.05.08.10.14.18.00.05.00.05.00.00.04.00.03.00.04.06.00.02
.21.25.08.06.06.11.11.04.05.00.03
.21.12.18.00.19.00.20.00.00.06.00.04
.20.10.12.09.11.08.06.00.08.12.00.00.04
.18.08.21.05.08.00.11.00.06.05.00.08      .04         .06
.45.10.10.10.05.20
```

Values of X^2 produced from the test of the random number generator (RANF-1) of the CDC 3600 Computer at the University of Massachusetts. These values represent a uniform distribution check for 130 replications of service time distributions containing 25 catagories (event types) each.

22.000	28.000	29.000
33.250	21.750	9.750
31.500	19.250	23.750
21.750	22.500	15.000
18.250	27.750	21.125
20.750	20.500	20.250
27.250	16.250	29.250
23.750	17.500	31.250
23.250	17.000	33.500
20.500	18.000	16.000
24.750	19.250	18.500
15.000	10.500	24.500
24.000	18.000	23.500
10.750	21.250	24.750
29.000	24.000	18.500
26.750	19.750	14.500
29.250	21.250	28.250
27.000	27.000	18.250
16.500	18.500	27.750
19.500	15.250	15.750
20.500	20.250	17.750
15.750	16.750	23.750
21.750	47.250 *	15.000
27.000	26.500	16.125
32.000	15.750	20.250
18.250	19.500	21.000
13.750	20.500	16.750
17.625	27.500	28.000
25.500	16.250	18.250
21.500	16.250	22.125
14.250	31.500	23.250
21.000	25.250	27.750
25.750	21.250	25.000
30.000	27.750	26.250
22.750	34.000	25.875
10.750	15.375	25.250
18.500	24.500	19.500
25.250	18.750	29.250
16.000	18.000	21.750
25.250	24.000	24.000
14.500	13.750	16.750
21.750	16.500	23.750
25.000	16.000	18.250
21.750		

* Significant at the .01 level. df= k-1=25-1=24