8-29-1986

An enrollment projection simulation

Dianne Bills

ROCHESTER INSTITUTE OF TECHNOLOGY
SCHOOL OF COMPUTER SCIENCE AND TECHNOLOGY

AN ENROLLMENT PROJECTION SIMULATION

by

Dianne P. Bills

A thesis, submitted to
the School of Computer Science and Technology,
in partial fulfillment of the requirements for
the degree of
Master of Science in Computer Science

Approved by: Name Illegible
Name Illegible
Name Illegible 10/1/86
Name Illegible

August 1986

An Enrollment Projection Simulation

CONTENTS

ABSTRACT

This thesis applies current computer science technology to the simulation of student behavior, called Student Flow, in higher education. Data Flow Diagrams were modified for Student Flow analysis and used to help select parameters relevant to the simulation. A longitudinal database of enrollment activity was retrieved and analyzed to generate simulation statistics. The simulation was implemented using an electronic spreadsheet. The model thus developed was then used to project future student enrollments.

CHAPTER 1

PROBLEM OVERVIEW

1.1 DATA FLOW DIAGRAM TECHNOLOGY

Structured Analysis is the first stage in the discipline of "software engineering". Structured Analysis uses "structured tools" in the analysis phase of system development to generate a model of a proposed system. Both DeMarco, and Gane and Sarson proposed the Data Flow Diagram (DFD), Data Dictionary, and Structured English (and/or Decision Tables and Decision Trees) as the structured tools needed to develop a "structured" system design[4,8]. A structured design produced with these tools is more accurate, requires less paper work, and is easier to understand and maintain than with non-structured techniques[5].

DFDs are dependency diagrams used in top-down analysis and design. As the first step in Structured Analysis, DFDs are used to graphically define basic system functions and their interfaces. They provide a concise and explicit graphical representation of logical system specifications. With careful balancing of inputs and outputs, the specifications produced are rigorous. However, they are also easy to modify and readily understood by users[4,5]. As illustrated by authors such as Weinberg, and Martin and McClure, and recent speakers, DFDs have become a standard systems analysis tool[1,13,20]. In the move towards automation of structured techniques, computerized versions of DFDs have been developed. EXCELERATOR by InTech [6] and STRADIS/DRAW by MCAUTO [16] are examples.

The major advantage of DFD technology is that it represents a system "from the viewpoint of the data"[4]. Thus, tracing data flow with DFDs, provides an integrated picture of an entire system. The literature cited above suggests, that for certain situations, DFDs are superior to other analytical techniques for the following reasons:

1. the system is defined strictly in terms of its inputs, outputs, and processes rather than by its procedural flow or data structures;

2. system inputs and outputs can be defined recursively;

3. when using DFDs to define a model, the rules of DFD construction (e.g. input/output balancing, and source/sink identification) add rigor and consistency.

The concept of "data" encompasses more than just computer readable input and output. Therefore, this basic characteristic of DFDs, the ability to graphically represent flow, would appear to make DFDs applicable to flow analysis in a variety of contexts. As noted by Martin and McClure, organizations tend to expand their use of DFDs beyond data and document flow[13]. One area that lends itself well to DFD analysis is enrollment projection in higher eduation. The behavior of individuals entering, progressing, and exiting an educational process is commonly called Student Flow[2,10]. Before student enrollments can be predicted for a given institution, Student Flow at that institution must be fully analyzed. One purpose of this thesis is to demonstrate the adaptation and effectiveness of the DFD technique to the analysis of Student Flow. Then, based on this flow analysis, a student retention simulation was developed which is used to project student enrollments.

1.2 SIMULATION IN OPERATIONS RESEARCH

Simulation is the process of using a mathematical model of the important functions of a system to study the effect of changes in environment and/or management policies on the real system. Simulation differs from other analytical modeling techniques because it has a higher degree of "isomorphism" to real life[9]. A series of events are generated, and the effects of the events on the system are chronologically traced. Simulation techniques are generally used for problems involving events over time that are too complex to be solved with other techniques. The advantage of simulation is its realism. The user can experiment with a model that resembles the real system, test scenarios, and measure results as if working with the actual system. The user can also quickly validate results with parallel simulations[9,14].

The reasons for choosing simulation are the same as for using any other modeling technique, to quickly and economically solve an operational problem. To do a simulation, the user needs:

1. data on the relevant attributes of each entity;

2. rules for the situations and decisions relating to each activity to be modeled;

3. the management policies concerning the problem[9].

The basis of predictive models in simulation is the "Input-Transformation-Output" or "Input-Process-Output" (I-P-O) model from general systems theory[3]. First a system is defined. Then logical relationships are developed for the important system functions. Together, these become the system model. Given input for the entities in each state of interest, the model will output a prediction of the dynamic system performance[3]. This model has been widely applied to Operations Research in higher education[15,17]. Of particular interest, Tinto used I-P-O as the basis of his longitudinal-process model of student attrition[17]. His model has become the most generally accepted and tested conceptual model in current attrition and retention work. It is diagrammed in Figure 1.

The model indicates that an individual's background characteristics determine the initial commitment to education and a particular educational institution. Educational experiences (grades, peer and faculty interactions, etc.) recursively affect educational and institutional commitment through the processes of academic and social integration. These levels of commitment should be the best predictors of student retention and attrition[15]. However, finding the factors that categorize students and parameters that accurately reflect these commitments is difficult[15].

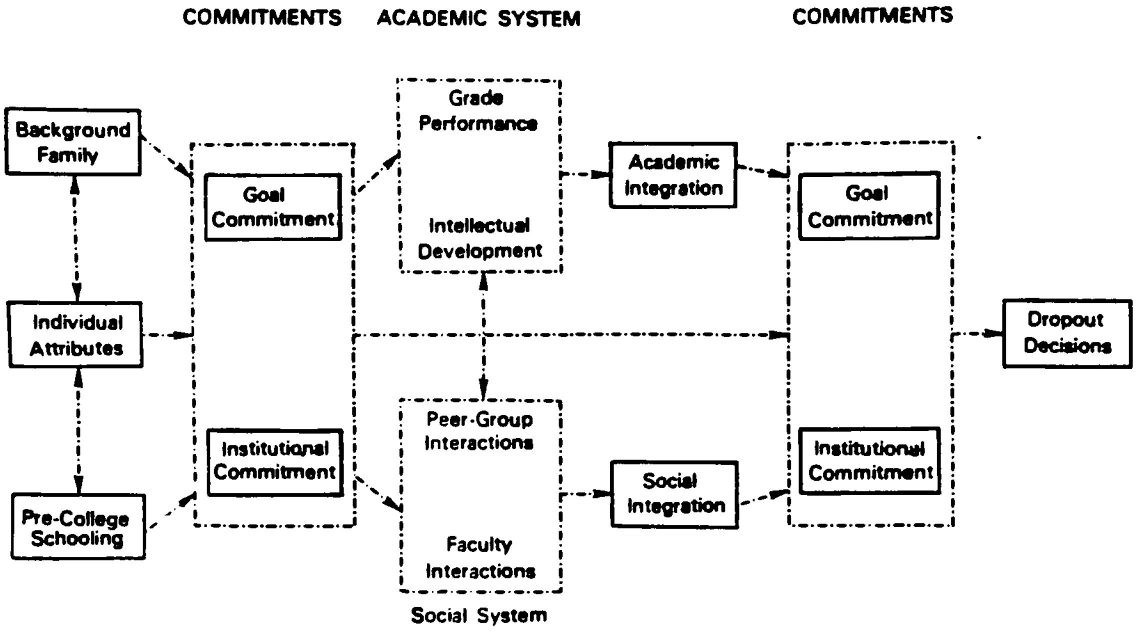


Figure 1. Tinto's (1975) Student Attrition Model

The Tinto model, and others, most notably Pascarella, provide a general conceptual framework for doing attrition and retention analysis[7,15]. However, they are by necessity high level models. To simulate retention, a researcher must understand what is happening and be able to identify relevant background, environmental, experiential, attitudinal, and outcome variables. This is a difficult task, made more difficult without a clear picture of the activities or "flows" within an educational system. Conceptual models, such as Tinto's, do not give the indepth picture of Student Flow that is necessary for accurate enrollment simulation, nor do they support the selection of simulation parameters.

Flow analysis is a key feature of DFDs. The flows within an educational system are the activities of students. Techniques which document procedures or data structures are not good matches to the needs of Student Flow analysis. Administrative policies and organizational or academic structure are not directly relevant to enrollment simulation, but student behavior is. Thus, DFDs are an excellent tool for analyzing attrition and retention in higher education. The DFD's ability to capture inputs and outputs can be used to help identify categories of students, such as freshmen, transfers, stop-outs, and drop-outs, etc. An indepth knowledge of student academic activity is required for enrollment modeling. With DFDs each category can be more easily and accurately followed throughout the educational process. DFDs can even help identify subgroups of students within categories that merit special analysis. DFD input and output balancing add rigor while supporting the documentation of important subgroups of students. Especially relevant is the ability of DFDs to handle recursion. Tinto's model indicates that academic and social integration in higher education are recursive processes. DFDs can be used to diagram this recursion.

DFDs can provide a detailed visual analysis of Student Flow. The conceptual models mentioned above do not. With a clear picture of the "flows" within an educational system, a researcher may be better able to identify the factors that affect and predict attrition and retention. DFD´s can be used to illuminate these important factors. Knowing these factors can help a researcher select and refine parameters relevant to a simulation.

The purpose of this thesis is to develop a student retention model which can be used to project future student enrollments for a programatically diverse institution of higher education such as the Rochester Institute of Technology (RIT). DFD technology is used to analyze Student Flow to find the important student activities which must be simulated by that model. Exactly how the standard DFD definition is "tuned" to analyze Student Flow is discussed in the next chapter. The next section discusses Student Flow and enrollment simulation in more detail.

1.3 ENROLLMENT SIMULATION

Enrollment projection is a simulation which estimates the size of the student pool for a given educational institution and the effect these potential students will have upon the institution. This information is of critical importance to an educational institution. The behavior of actual and potential students determines the viability of individual programs, and ultimately the viability of the entire institution. Enrollment modeling and projection are part of the total enrollment management process of an educational institution. Enrollment management has two aspects, external and internal to the institution. External enrollment modeling is part of the recruitment process. It defines the size of the institution's potential student pool, and estimates the number of applications and actual admission yield for each program of study. Internal enrollment modeling deals with student attrition and retention. The behavior of students matriculating in each program of study is analyzed. In the past, enrollment modeling in higher education concentrated mainly on recruitment. However, as the size of the available student pool has decreased, competition between institutions of higher education has increased[10,11]. As discussed in the next chapter, recent simulation focuses on factors that define how long students stay in the educational environment.

Wasik used regression and Markov Chain theory to define a Student Flow model for institutional enrollment as follows[19]:

GIVEN: total new enrollment = E = f(H, M, A, R)

WHERE: H = numbers of high school graduates by year;
M = military manpower needs by year;
A = an estimate of economic activity by year;
R = recruitment area population.

LET: Q_t = total institute enrollment at time t;

P_t = a transition matrix whose [i,j]-th entry is the probability that a student in program i at time t will be in program j at time t+1;

C_t = a column vector of proportions of students in various programs averaged over a fixed number of years;

E_t = total new enrollments at year t;

D_t = a column vector of proportions of new students who enroll in various programs averaged over a fixed number of years.

THEN: the distribution of students expected to be enrolled in each program at time t $= X_t = Q_t C_t$;

and the vector of frequencies of new students expected to be enrolled in various programs at time t $= N_t = E_t D_t$,

THUS, total institutional enrollment at time t+1 $= F_{t+1} = X_t^T P_t + N_t$.

PROBLEM OVERVIEW

Markov models are linear fractional flow models in which the current state is assumed to depend only upon the previous state, ignoring history[12]. However, it is possible that the enrollment distribution vector X_t, may actually depend upon more than just the preceding enrollment state. Thus, the capital "T", in the final formula above. It indicates that a student´s enrollment state at time t may actually be based upon the pattern of two or more past enrollment states, T, rather than just the previous one.

This model clearly illustrates both the internal and external components of Student Flow. It also shows the recursive aspect of Student Flow that makes DFD analysis so applicable. It is included solely for those reasons. As noted by Wasik, single-state dependency can not be applied to the educational process without accepting over-simplification. Markov models are popular because they are conceptually simple, and can be used where resources and data are limited[19]. More relevant to the current analysis, however, Markov transition probability matrices apply to best attrition analysis. This study projects future student enrollments by simulating student retention. It uses a modification of Cohort Survival Methods. This is discussed further in the implementation chapter.

PROBLEM OVERVIEW

The next section highlights specific enrollment issues and problems at RIT that initiated this study.

1.4 ENROLLMENT PROJECTION AT RIT

The RIT Enrollment Projection Process is currently handled by the Office of Institutional Research under the Division of Institutional Advancement.

The RIT Enrollment Projection Process was designed to give the Institute a "picture" of the characteristics and enrollment potential of each college for the upcoming year plus aid in resource allocation. The projections support the RIT budget process by supplying enrollment data to RIT's Budget Officer and the Administrative Committee. They also provide feedback to the academic deans and Academic Affairs through the Academic Planning Process. The current Enrollment Projection Process combines historical enrollment data with information on student availability from RIT Admissions.

RIT currently has a biannual enrollment projection process. Using historical data from the past three (3) years, enrollment projections for the upcoming year are done in the fall quarter of each academic year. These projections are shared with the deans of each college through the Academic Planning Process. At these meetings the size of the available student pool and any policies or recruiting strategies that will affect enrollment are discussed and the projections are adjusted where appropriate. The "final" projections from fall quarter are reviewed and further adjusted in the following spring quarter.

The primary data supplied by the current Enrollment Projection Process are estimates of student enrollment for the upcoming year by college, department, and major. The level of detail at which projections are done varies with the college. Student headcounts are projected by year in program for the categories of: full-time students, part-time students, and coop students. Other important measures such as FTE counts, estimates of student quality, student/faculty ratios, and instructional costs/credit hour are derived from the enrollment projections.

The Office of Institutional Research at RIT has "grown into" the current enrollment projection process over a number of years. It is now becoming apparent that the process itself has become unwieldy. The enrollment projections produced while generally felt to be "good", are now clearly not good enough. RIT is for the first time entering a period of declining enrollment. A mathematically sound model (or group of models) is needed. RIT must accurately project the unique situations in each college so that appropriate planning can occur.

PROBLEM OVERVIEW

The areas of concern with the current RIT projection process are:

1. The current procedure for generating enrollment projections involves a cumbersome, difficult, and error-prone manual process. Many extra man-hours are required to produce projections, and it is extremely difficult to catch transcription errors that get into the data.

2. After the enrollment projections are reviewed with each college, they are further adjusted. Currently, this means wading through the same cumbersome manual process that is used to produce the projections. An easy and flexible method of modifying the original projections is needed.

3. The enrollment projection model itself needs to be improved. The current algorithm is a straight three year enrollment average. A more realistic algorithm is needed.

4. The current projection process only produces projections for the next academic year. Increasing financial pressures make a model that will project for multiple years into the future imperative.

5. The projection system must be less "person dependent". It should be easy to use and clearly documented so that someone new can be trained with minimal effort.

6. Finally, as discussed by Pascarella, the conceptual models of Tinto and others are intended for longitudinal analysis[15]. The advantages of longitudinal analysis over other data storage designs for simulation are widely accepted[15]. Unfortunately, RIT enrollment data is stored cross-sectionally. A longitudinal database design is needed.

CHAPTER 2

DATABASE DESIGN

2.1 ANALYZING STUDENT FLOW

As mentioned earlier, Student Flow is the popular term for the movement of individuals through their course of study at an institution of higher education. Before this process can be simulated, its "flow" must be analyzed to find the controlling factors and pertinent parameters.

Student Flow may not appear to lend itself to structured analysis techniques used in Computer Science. This "flow" is composed of the educational activities of unique individuals not data. Plus, each person's educational plan can, potentially, be unique. However, the educational choices available at a particular institution will tend to group students into natural categories. Individuals will be grouped with other individuals who have similar educational plan and progress characteristics. The movement of these groups of students can be modeled based on the characteristics, or parameters, that define their progress. Individual student parameters can be collapsed and analyzed across groups of students to find the underlying factors that model Student Flow. While various analytical tools are available, the DFDs seem to be the natural choice for identifying the factors which characterize Student Flow.

DFDs trace the flow of data through its use in an organization by diagraming functions and interfaces. Bubbles represent unique processes that use or transform the data. Arrows show the flow of data items into and out of each process. DFDs provide an unbiased picture of all possible "paths" in the system. Thus, as mentioned by DeMarco, they are a convenient tool for modeling real life situations[4]. Just as DFDs document data flow, they can be used to help visualize Student Flow. DFDs supply conceptual documentation, modeling, and automatic system partitioning. These are the framework for creative analytical thought that are missing from models like Tinto´s.

The "data" of education are students. By adjusting DFD technology to represent the activities of individuals in higher education, DFDs can be used to analyze Student Flow. Let each bubble represent some enrollment filter. An enrollment filter is defined to be a process within the educational environment that causes students, or potential students, to terminate their involvement with an educational institution. Arrows normally represent "pipelines" for "packets of information"[4]. The packets will now be one or more individuals continuing-in or terminating their educational plans. With the arrows representing groups of individuals, procedural annotation will not be used because exclusive conjunctive or disjunctive situations are impossible. Although in standard DFDs, processes can create new data items, an enrollment filter will not be allowed to "create" new students. With these redefinitions, DFDs can be used for Student Flow simulation analysis.

The fundamental DFD for the Student Flow process is in Figure 2, below.

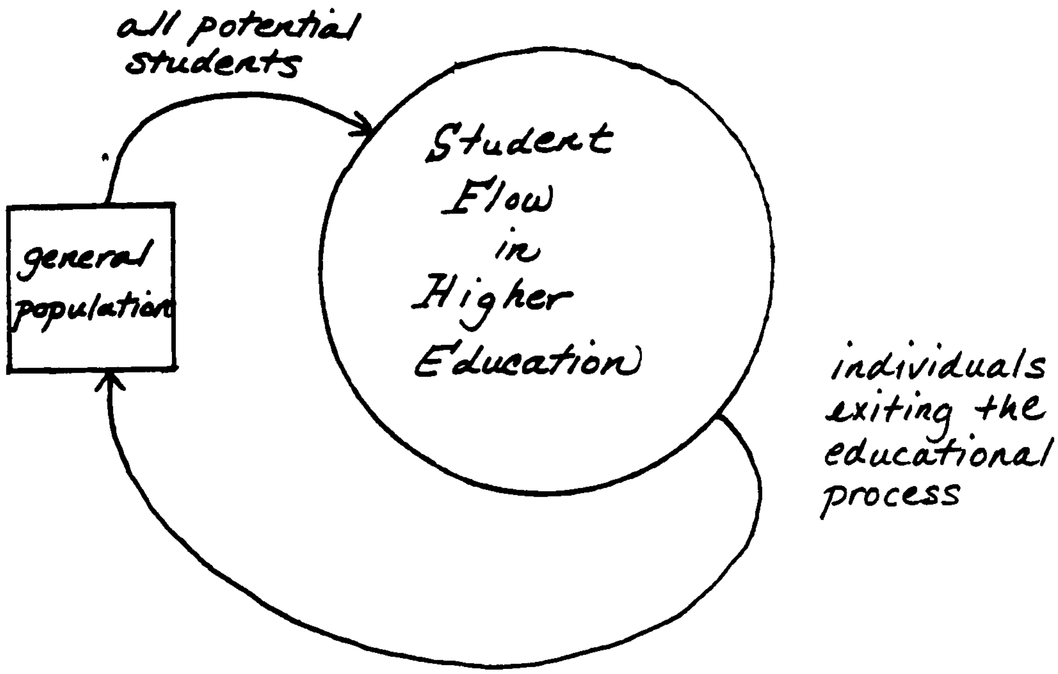


FIGURE 2. STUDENT FLOW FUNDAMENTAL MODEL

The previous diagram can be expanded as in Figure 3.

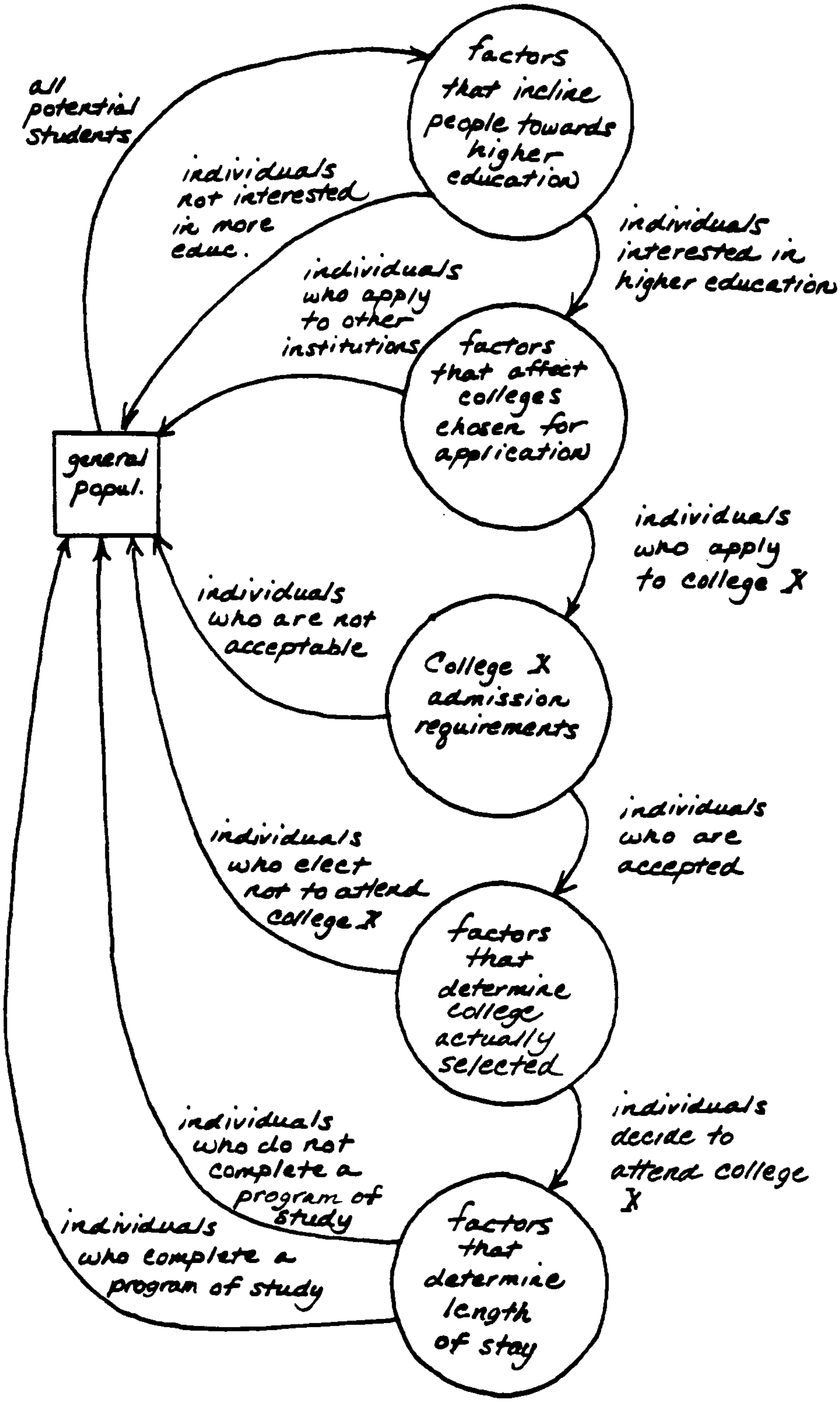


FIGURE 3. STUDENT FLOW IN HIGHER EDUCATION

The ultimate source, and sink, is the general population. The primary input is all individuals who, potentially, are interested in higher education. The primary output is individuals who are exiting the educational process. This output encompases two groups, individuals who complete their studies and individuals who do not. It should be noted that both these groups can reenter the educational process as potential students at a later point in time.

The enrollment filters are the various factors that incline individuals to want higher education, pick a particular educational institution, apply and be admitted, and finally exit a chosen course of study. The term "program of study" is used to avoid the traditional connotations associated with the terms "graduate" and "dropout". The current thinking is that each individual´s personal goals and skills determine the "success" of an educational program. Therefore, academic success should not be measured solely by numbers of certifications[15].

The focus of this analysis is the last filter, "factors that determine length of stay". It is further refined on the next page. This last filter is of major interest because of the shrinking student pool in higher education. As mentioned previously, the projected decrease in higher education students through 1990, has caused greater competition between institutions. Thus,

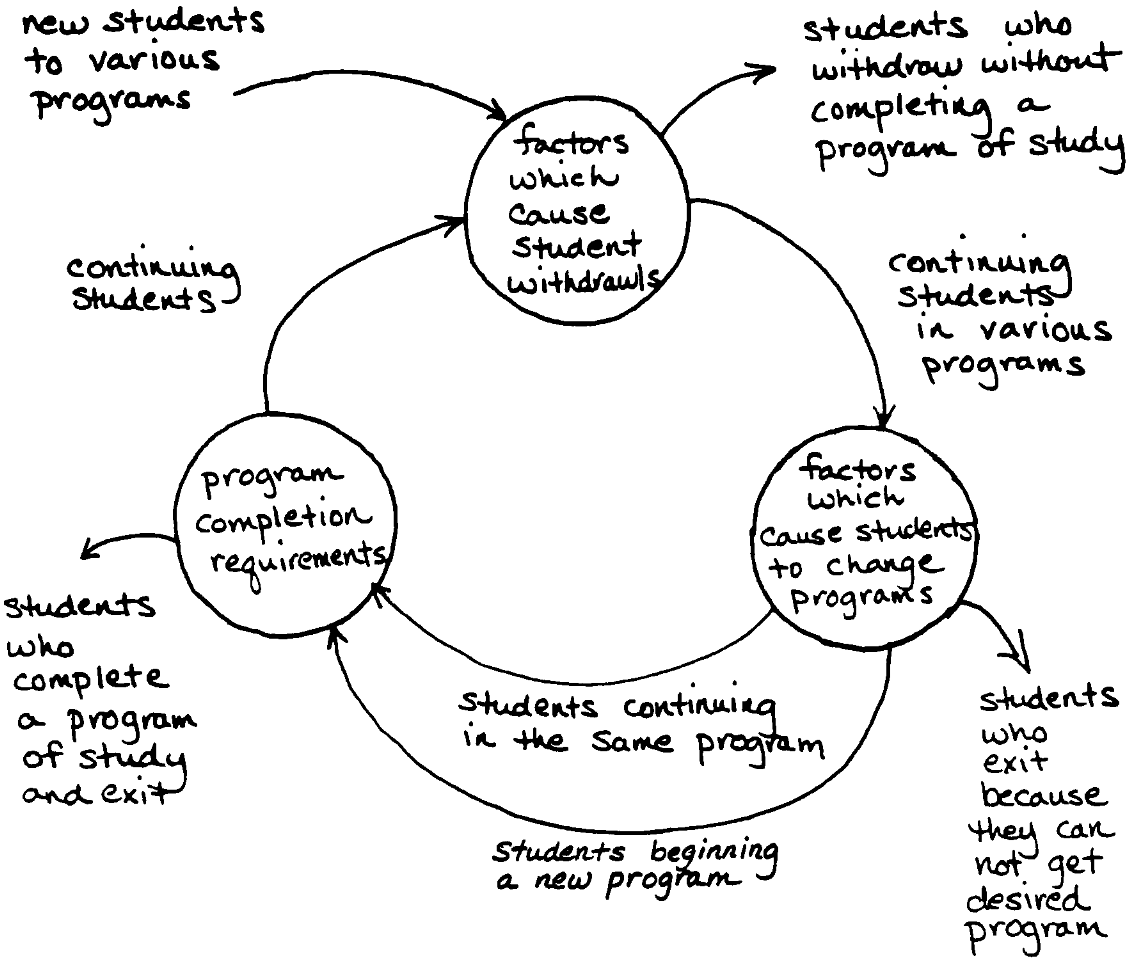


FIGURE 4. FACTORS THAT DETERMINE LENGTH OF STAY

recent research in enrollment modeling has centered on the factors that determine whether or not a student actually completes his desired program of study. As illustrated by Walter, interest has increased in analysis of factors that influence length of stay in an educational environment as a means of maintaining enrollment levels[18]. Therefore, this model concentrates on Student Flow after students have been admitted to an institution of higher education.

The DFD refinement above shows the categories of factors that influence length of stay. The weight of each of these factors can vary with program choice. Notice that while DFDs are technically independent of time, this diagram has an inherent time factor. This is because the educational process occurs incrementally over successive time intervals. Pressures to withdraw or change program can affect individuals repeatedly and with variable weight throughout their course of study. For example, the difficulty of maintaining a passing grade point average can vary with the courses taken. And students are more likely to transfer to another program early in their course of study rather than later when they are closer to certification. Thus, the recursive nature of these last filters.

The model developed in this study simulates student retention to project future enrollments. For retention simulation, it is only important to know whether students continue to be enrolled at or exit an institution. So, whether individuals exit because they complete their program of study, or decide not to complete it is not important. Thus, all exiters can be combined into one category as shown in Figure 5.

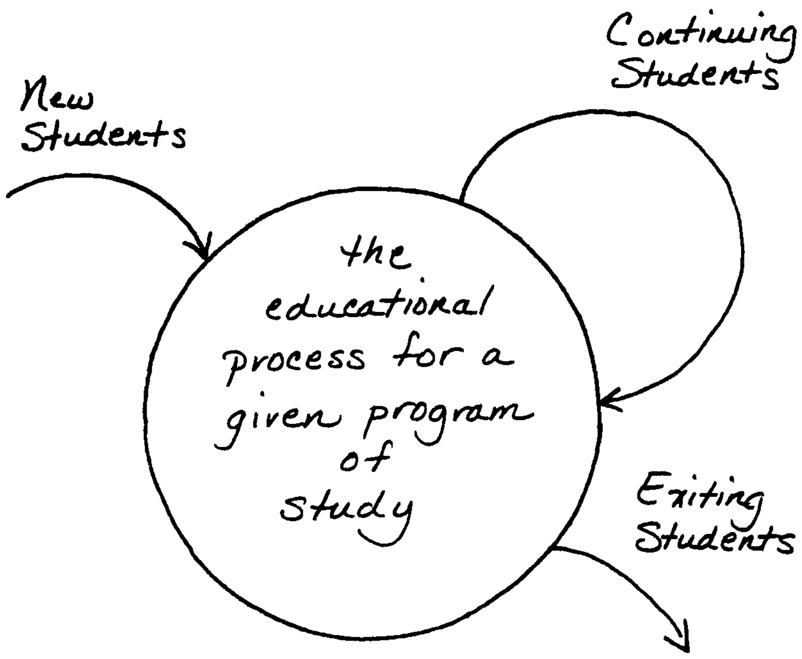


FIGURE 5. STUDENT FLOW RETENTION ANALYSIS ACTIVITIES

This final DFD is the "kernel" of this Student Flow analysis. It shows that to simulate retention for enrollment projection, whether for one program or an entire institution, a model must account for the student categories of new admits, continuing students, and exiting students.

As discussed at the beginning of this chapter, common characteristics will tend to define students into groups with similar flow rates. Therefore, to model the student activities shown in the final DFD in Figure 5, parameter categories must be chosen which define the important flow groups of each student category identified. Based upon the current literature, these parameter categories are[15,18]:

1. the program of study an individual chooses[15]. This field can be used to select students corresponding to the level of detail desired for the simulation.

2. the student´s "entry status", parameters that define a student´s previous educational experiences. This is necessary because students with different educational background tend to progress at different rates[18].

3. the student´s "progress rate", parameters that measure the rate at which an individual moves towards completing his or her chosen program of study[18].

4. and from the DFD in Figure 5, student "status", whether an individual is newly admitted to, continuing in, or exiting the program area under analysis.

Taken together, program choice, entry status, progress rate, and student statuses are the arguments of a four dimensional model for projecting student enrollment in a program of study at an institution. This four dimensional model is diagrammed in Figure 6 below. Note that the relationship between the arguments may not be orthogonal. Although the diagram is drawn assuming there is no correlation between the arguments, they may not be fully independent. For example, some college programs only accept transfer students who have credits equivalent to an AAS.

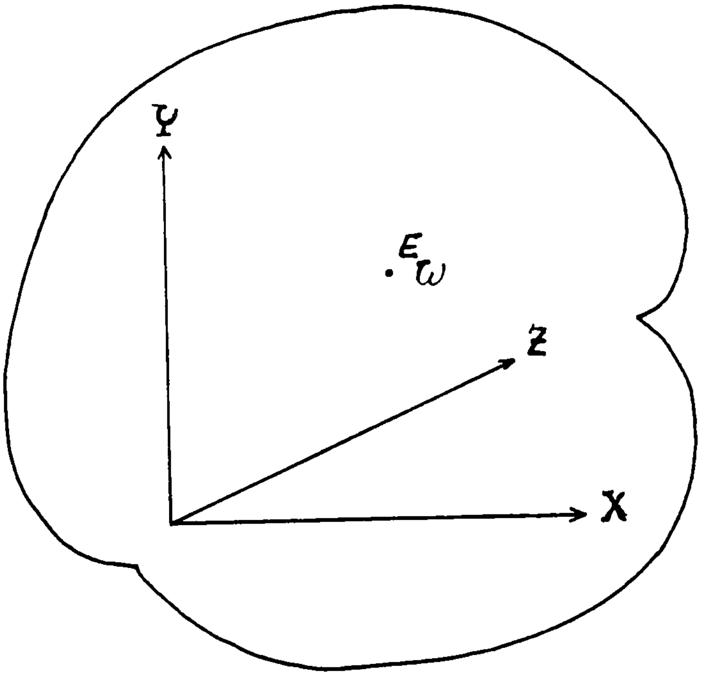


FIGURE 6. ENROLLMENT RATE FOR PROGRAM CHOICE "W"

As shown in Figure 6 above, the future enrollment size of a given academic program W, E_W, is function of the previous educational background, X, length of stay or progress rate, Y, and current academic activities, Z, of its matriculants. Total enrollment for an educational institution at a given point in the future can be estimated as the sum of the enrollments of each individual academic program projected for that time.

For the purposes of this project, DFDs have been used strictly as a conceptual tool to aid in problem definition. Specifically, they have been used to analyze Student Flow, and identify the important factors and student activities which a retention simulation must handle. DFDs are not used to define actual simulation logic or program functions. Now that the important student activities have been identified, retention simulation and enrollment projection are mathematically derivable from the parameters listed above.

The next step in the development of a system model using Structured Analysis tools is the creation of a data dictionary. For Student Flow analysis, this means choosing actual data items which represent each of the important student categories identified in the flow analysis. These data items are used to create a model of Student Flow. The next section discusses the actual data elements and database structure necessary for this simulation.

2.2 THE ENROLLMENT PROJECTION DATABASES

The previous section identified four parameter categories that can be used to define flow groups and the rates of flow for students at an institution of higher education. When modeling Student Flow, these parameters appear in two different databases. The first is a detail level database called the Student History Database. It contains individual student historical data with identification parameters in addition to those in the four parameter categories. The second database is a Rates of Flow database. Collapsed historical data for each academic area are loaded into separate Rates of Flow databases for analysis. Student Flow simulations for a given academic area are done from this database. While both databases have parameters in each of the four categories, the Rates of Flow database generally contains the current "working set" variables. The Student History database may contain additional data items in each category not in the current working set as well as demographic data. These data items can support creative investigation of the various rates of flow.

To simulate Student Flow, Rates of Flow databases for each student group to be modeled are generated from historical student information between two fixed points in time. Most commonly, data for three to five years back, measured at each successive Fall, is used. The actual variables necessary for modeling Student Flow can vary depending on the characteristics of an institution. However, the DFD analysis in the previous chapter indicates the following as generally applicable:

THE RATES OF FLOW DATABASE

PARAMETER	DESCRIPTION
1. Program of Study	A code for the student´s program.
2. "Entry Status"	A parameter which defines students into groups based on their previous academic experience: • students who have no previous college experience, called "New Students" • students who tranfer in with some previous college experience, called "External Transfers" • students who transfer in from another program within the same institution, called "Internal Transfers"

THE RATES OF FLOW DATABASE (con´t.)

PARAMETER	DESCRIPTION
3. "Progress Rate"	A measure of the student´s progress towards completing his program of study.
4. "Current Status"	A status variable which categorizes students as follows: • students who are continuing in a program of study • students who are new to a program of study • students who have exited the program of study

The parameters above define a three dimensional "table" with the proportions of students in each of the "Current Status" categories for one program at one point in time. This table is a transition matrix for that point in time. A Rates of Flow Database for a given program combines transition matrix data across all points in time measured. The proportions of students in each entry, progress, and status category combination across the chosen points in time are the rates of flow for students in a particular program.

Data from the Student History Database is used to generate individual Rates of Flow databases. It contains the longitudinal student detail data necessary for Student Flow modeling. Rates of Flow databases can be generated at the institute, college, department or program level depending upon the needs of the modeler. The variables above are collapsed across selected students to setup the statistics needed for the Rates of Flow Databases. As previously mentioned, the Student History Database may contain other data items which can be used for flexible analysis and investigative modeling capabilities. The layout of the Student History Database used for this simulation is in Appendix B.

The Rates of Flow Database for this simulation is stored in the 20/20 spreadsheet used to implement this simulation. The next chapter discusses the use of spreadsheets for modeling, and the process used to generate the Student History and Rates of Flow Databases for this particular simulation.

CHAPTER 3

STUDENT FLOW MODEL IMPLEMENTATION

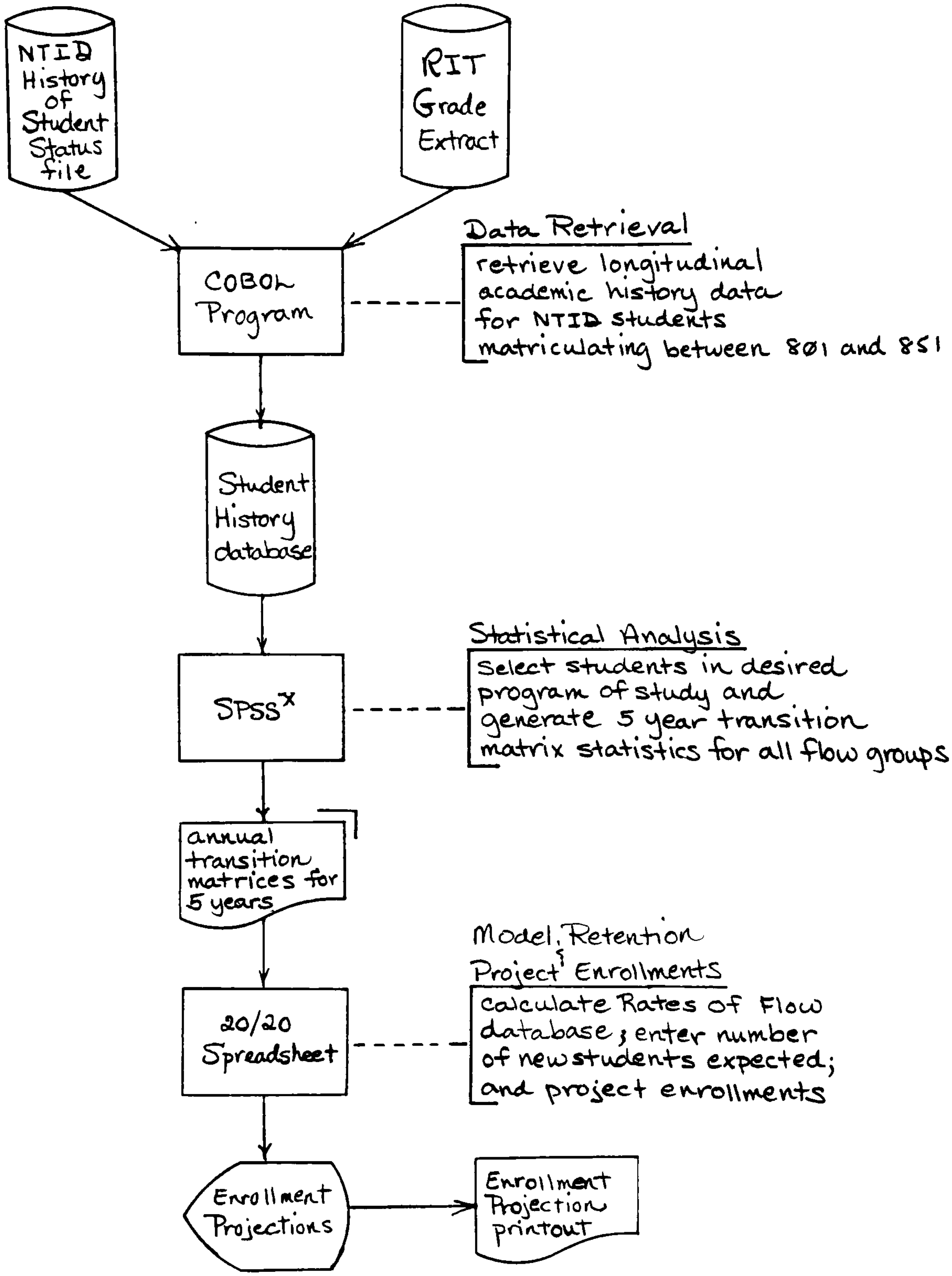
Electronic spreadsheets are a recent simulation tool. An integrated spreadsheet with statistical analysis functions and graphics will support flexible modeling for a limitless variety of applications. The main advantage of electronic spreadsheets is that model parameters and assumptions can be changed and a computer recalculates the simulation results. Time and effort are greatly reduced. Thus, planning and decision making are easier[21].

A system flowchart of the data retrieval, analysis, and simulation process is on the next page. The process is described below.

3.1 DATA RETRIEVAL

For this enrollment projection and student retention simulation, longitudinal Student Flow data for the past

FIGURE 7. THE ANALYSIS AND PROJECTION GENERATION PROCESS



five years measured at each Fall quarter was used. The sample of data analyzed is for NTID-supported students active between Fall Quarter 1980-81 (801) and Fall Quarter 1985-86 (851). This data was retrieved from the NTID History of Student Status and the RIT Grade Extract files with a COBOL program. The data retrieved included original RIT entry status and current academic status for all students. In addition, quarterly academic status, program code, and credit hours earned were retrieved for each quarter of the six year period. Refer to Appendix B for the complete file definition. The retrieved historical data was analyzed using the SPSS-X Statistical Package.

3.2 STATISTICAL ANALYSIS

SPSS-X was used to categorize students into flow groups and generate the transition matrix statistics required for the simulation. The following student characteristics, previously identified in the DFD analysis, were used to assign students to flow groups for analysis:

1. the students' program of study at RIT/NTID;

2. the students' previous college experience prior to admission to the program of interest;

3. the students' year in program as a measure of progress towards completion of a program of study, and

4. the students' current status in the program of interest.

Program of study is the standard RIT four character code. The previous college experience, or entry status, categories are the following:

1. new to college (no previous college experience);

2. external transfer from another college or program;

3. internal transfer from another program within RIT/NTID.

These categories were assigned based upon the students' most recent experience prior to entering the program under analysis.

Progress towards completion of a program was based upon a calculated year in program. Year in program was estimated from the credit hours completed in the program of interest up to the point of measurement. It is an integer between one and seven inclusive. Half credit was given for all credit hours accumulated while not in the program of interest. Year in program was calculated assuming the average student completes thirty-six credits towards a degree in one year (three quarters at twelve credits per quarter).

At each Fall Quarter measured, students were recategorized into one of the following "current" status categories based upon their status in the base (previous Fall) term:

1. actively matriculating in the program of interest in both the base and measure (Fall) terms;

2. a new admit to the program after the base term and in or before the measure term;

3. exited the program (either graduated or withdrawn) after the base term and in or before the measure term.

These three entry, seven progress, and three status categories separate students into sixty-three categories for the program under analysis at each measure term included in the analysis.

The SPSS-X program was used to select students in one academic program, NTID Data Processing (NBTD), for analysis. This program also contained the logic to assign each student to an entry, progress, and status category based upon his or her individual academic history. This three way assignment was repeated for each measure term. Cross-tabulations of the numbers of students in each of the sixty-three possible academic states were done for each of the five base and measure term intervals across the six years of data. For example, the students active between Fall 801 and Fall 811 were tabulated into the three entry by seven year in program combinations as follows:

1. the numbers of students in each of the twenty-one categories who exited the program after 801 but in or before 811 by year in program at 801;

2. the numbers of students in each category who were actively matriculating in both 801 and 811 by year in program at 811;

3. and the number of new admits to each category after 801 and in or before 811 by year in program at 811.

This same analysis was repeated for 811 to 821, 821 to 831, 831 to 841, and finally 841 to 851. These numbers for these five year intervals were then entered into a 20/20 Spreadsheet for analysis.

3.3 MODEL RETENTION AND PROJECT ENROLLMENTS

The 20/20 Integrated Spreadsheet Modeling Program was used to store the five year retention history data and implement the enrollment projection simulation. The enrollment projection model developed here is a modification of Cohort Survival Methods. The standard Cohort Survival technique calculates successive survival rates for a given group of individuals across one or more points in time[12]. It is modified here by adding in the number of new students admitted to the program of interest at each Fall term. This modification is necessary to support the calculation of annual continuing student enrollments needed for the enrollment projections.

As mentioned earlier, student counts in each of the sixty-three categories, for each of the five base to measure year intervals, were entered into the 20/20 Spreadsheet. Based upon these figures, the spreadsheet was used to calculate flow statistics for each student category: exiting, continuing, and new. The following flow statistics were calculated:

1. the proportions of students continuing to be enrolled in each of the entry and progress categories at each measure term (Fall quarter);

2. the proportions of new students admitted into the entry and progress categories between each measure term;

3. the proportions of students who exit from each entry and progress category between each measure term.

These three vectors calculated at each Fall Quarter are a transition matrix for that measure term. The transition matrix for (Fall 801 to) Fall 811 as stored in 20/20 is below.

ENROLLMENT PROJECTION MODEL -FOR: NTID DATA PROCESSING

		YEAR 0-1		YEAR 1			
		EXITING STUDENTS		CONTINUINO STUDENTS		NEW STUDENTS	
YEAR IN PROGRAM	ENTRY STATUS	N	.P	N	.P	N	.P
1	NEW	20	0.400	2	0.028	35	0.530
	EXT. TR	0	0.000	1	0.014	17	0.258
	INT. TR	3	0.600	0	0.000	10	0.152
2	NEW	12	0.387	29	0.408	0	0.000
	EXT. TR	0	0.000	4	0.056	0	0.000
	INT. TR	4	0.333	3	0.042	4	0.061
3	NEW	10	0.667	17	0.239	0	0.000
	EXT. TR	5	0.833	0	0.000	0	0.000
	INT. TR	1	1.000	6	0.085	0	0.000
4	NEW	5	0.833	6	0.085	0	0.000
	EXT. TR	0	0.000	1	0.014	0	0.000
	INT. TR	0	0.000	1	0.014	0	0.000
5	NEW	1	1.000	1	0.014	0	0.000
	EXT. TR	0	0.000	0	0.000	0	0.000
	INT. TR	0	0.000	0	0.000	0	0.000
6	NEW	0	0.000	0	0.000	0	0.000
	EXT. TR	0	0.000	0	0.000	0	0.000
	INT. TR	0	0.000	0	0.000	0	0.000
7	NEW	0	0.000	0	0.000	0	0.000
	EXT. TR	0	0.000	0	0.000	0	0.000
	INT. TR	0	0.000	0	0.000	0	0.000
TOTAL STUDENTS		61	0.462	71	0.538	66	0.482

FIGURE 8. THE FALL 811 STUDENT TRANSITION MATRIX

The annual vector proportions across the five measure quarters are the rates of flow. Rates of flow were calculated for the three student categories: exiting, continuing, and new. The spreadsheet was used to generate the following rates of flow statistics for each of the twenty-one entry by progress rate categories in the three student status categories:

1. individual "continue rates" for each entry and progress category;

2. the rate and direction in which each "continue rate" changed;

3. the overall rate at which students "continued" to matriculate;

4. the rates at which new students were admitted to each of the entry and progress categories;

5. the rate and direction of change of each individual "new admit" rate;

6. the overall annual proportion of new students;

7. the individual "exit rates" for each entry and progress category;

8. the rate and direction in which the "exit rates" changed;

9. The overall rate at which students exited the program or area.

Together, these calculations constitute a Rates of Flow Database for the program under analysis. The rate and direction of change for the proportions of new, continuing, and exiting students are trend factors. This database was used to generate the enrollment projections. The Rates of Flow Database for NBTD for Fall 1981-82 through Fall 1985-86 calculated for this simulation appears on the next page.

To generate projections from this model for a given academic program, the following inputs are needed:

1. the number of new students expected to enter the program at each future time to be projected;

2. and the rates of flow statistics from the Rates of Flow Database for that program.

The model projects the number of new and continuing enrollees by simulating internal Student Flow for one or more points in the future. Enrollment is projected for each of the entry status and progress rate categories for new and continuing students based upon history. The trend factor for each group is included in the calculations.

YEAR IN PROGRAM	ENTRY STATUS	AVERAGE EXITING YEARS 1-5	AVERAGE CONTINUING YEARS 1-5	AVERAGE NEW YEARS 1-5	AVERAGE CHANGE FOR EXITING YEARS 1-5	AVERAGE CHANGE CONTINUING YEARS 1-5	AVERAGE CHANGE FOR NEW YEARS 1-5
		.P	.P	.P	.P	.P	.P
1	NEW	0.338	0.039	0.619	-0.0464	0.0047	0.0213
	EXT. TR	0.149	0.013	0.134	0.0357	-0.0035	-0.0144
	INT. TR	0.310	0.010	0.156	-0.0964	0.0029	-0.0225
2	NEW	0.233	0.331	0.008	-0.0726	-0.0246	0.0000
	EXT. TR	0.225	0.064	0.000	0.0357	-0.0053	0.0000
	INT. TR	0.441	0.101	0.060	-0.0064	0.0172	0.0079
3	NEW	0.543	0.210	0.002	-0.0088	0.0103	0.0000
	EXT. TR	0.419	0.036	0.000	-0.0833	0.0073	0.0000
	INT. TR	0.697	0.058	0.009	-0.0556	-0.0080	0.0038
4	NEW	0.556	0.073	0.000	-0.0655	-0.0124	0.0000
	EXT. TR	0.267	0.013	0.000	0.0833	0.0009	0.0000
	INT. TR	0.250	0.018	0.012	0.0000	0.0009	0.0038
5	NEW	0.600	0.024	0.000	0.0000	0.0038	0.0000
	EXT. TR	0.000	0.000	0.000	0.0000	0.0000	0.0000
	INT. TR	0.000	0.005	0.000	0.0000	0.0058	0.0000
6	NEW	0.000	0.002	0.000	0.0000	0.0000	0.0000
	EXT. TR	0.000	0.000	0.000	0.0000	0.0000	0.0000
	INT. TR	0.000	0.000	0.000	0.0000	0.0000	0.0000
7	NEW	0.200	0.002	0.000	0.2500	0.0000	0.0000
	EXT. TR	0.000	0.000	0.000	0.0000	0.0000	0.0000
	INT. TR	0.000	0.000	0.000	0.0000	0.0000	0.0000
OVERALL AVERAGE		0.355	0.612	0.451	-0.0499	0.0467	-0.0516

FIGURE 9. THE NBTD RATES OF FLOW DATABASE FOR 811 TO 851

New and continuing student projections in each category are summed to produce the total numbers of new and continuing, and the total class size for a given year. The actual enrollment projections for NBTD for Fall 1986-87 (current year + 1) are below.

PROJECTIONS:

YEAR IN PROGRAM	ENTRY STATUS	YEAR +1 CONTINUING STUDENTS (N)	YEAR +1 NEW STUDENTS (N)	YEAR +1 TOTAL STUDENTS (N)
1	NEW	7	45	52
	EXT. TR	1	8	10
	INT. TR	2	9	11
2	NEW	47	1	47
	EXT. TR	9	0	9
	INT. TR	18	5	23
3	NEW	34	0	34
	EXT. TR	7	0	7
	INT. TR	8	1	9
4	NEW	9	0	9
	EXT. TR	2	0	2
	INT. TR	3	1	4
5	NEW	4	0	4
	EXT. TR	0	0	0
	INT. TR	2	0	2
6	NEW	0	0	0
	EXT. TR	0	0	0
	INT. TR	0	0	0
7	NEW	0	0	0
	EXT. TR	0	0	0
	INT. TR	0	0	0
	TOTAL STUDENTS	152	70	222

FIGURE 10. NBTD ENROLLMENT PROJECTION FOR FALL 86-87

The model also includes a set of weighting factors for each of the five years. The effect of new, continuing, and exiting student histories can be weighted as desired. Weights can be used to increase or decrease the contribution of certain groups or years on the projected results. For the projections done for this project, all years and student statuses were weighted equally.

The complete 20/20 spreadsheet with the transition matrices, Rates of Flow Database, and projections done for this thesis is attached in Appendix C. All calculations and formulas used in the spreadsheet are detailed in Appendix D.

3.4 MODEL RESTRICTIONS

The restrictions to the model developed are as follows:

1. The model simulates internal Student Flow only. It does not project the size of the external student pool from which the institution can admit students. Admission figures for each term are an input to the model.

2. This model simulates enrollment flow and retention rather than attrition. Therefore, no distinction is made between individuals who do or do not complete a program of study. Both are grouped together as "exiters".

3. The projections are based upon the activities of students actively matriculating in their programs of study at the points in time measured. Students who were temporarily inactive or on suspension during any Fall Quarter included in the analysis were intentionally excluded from the analysis.

4. The model projects numbers of internal and external transfers based upon the history of a given program or area. However, these parameters

require careful management by the person responsable for the projections. Other information from admissions or academic departments on student availability and new programs can significantly impact projection validity.

5. The program being modeled must have sufficient history from which to develop the Rates of Flow Database. If there is not sufficient past history, then suitable estimates must be available.

6. Finally, as with any simulation, there is a practical limit to how far into the future accurate projections can be done. Projected outcomes can not be viewed in a vacuum. Beyond a certain point, the probability that external or internal factors will override history becomes too large to depend soley upon mathematical projections.

The final chapter discusses the results of this simulation, the usefulness of spreadsheets, and the applicability of DFD technology to Student Flow simulation.

CHAPTER 4

RESULTS AND SUMMATION

Analyzing Student Flow using DFD technology was helpful in developing a model to project student enrollments. Clear understanding of a process is essential before it can be modeled. DFDs support clear visualization of a process on paper. In the case of Student Flow, DFDs bring out the categories of continuing, new, and exiting students which must be accounted for in any enrollment simulation. If there is any conflict between DFDs and Student Flow analysis, it is that DFDs are by definition time independent. Student Flow, however, is a process measured over time. In the final two detail level DFDs in chapter two it becomes necessary to quantify time as sucessive quarters of the educational process in order to clearly see the actions of individual students. This does not appear to decrease the effectiveness of DFDs in this application. For DFDs, the important characteristics are the graphical representation of flow and balanced leveling. These functions make DFDs applicable to almost any type of flow analysis.

Electronic spreadsheets are an excellent tool for Student Flow modeling and enrollment projection. The spreadsheet displays each step of the process. Intermediary calculations are available for verification, and each step of the modeling process is open to visual review. Thus, errors are less likely to become buried within internal logic as with traditional programing techniques. Also, spreadsheets allow quick recalculation. Various scenarios and "what if" questions can be quickly and easily tested. So flexibility and productivity can be dramatically increased.

From a staff training and ease of use standpoint, enrollment projection with a spreadsheet is a good choice. Spreadsheets are easy for users with a variety of computer skills to learn. Manipulation and management are simple and quick. To do projections with the model developed here, the user only enters new, continuing, and exiting student counts for the past six years. This data should be available in some form at every institution. The input data can even be "imported" directly from a computer file to save typing. Plus, this model can be used at a variety of levels, from individual programs and departments up to an entire institute.

RESULTS AND SUMMATION

In terms of enrollment management, individual program level projections as done here are probably the most useful. Projections for individual programs can always be combined to project upward for a department, school, college, and institute. And changes in individual program enrollments most directly affect course requests and staffing. Projecting at a higher level, such as department, may provide accurate overall numbers, but hide important program-level changes and trends.

Aside from the yearly projections, the most important calculations in the 20/20 spreadsheet are the average rates of change for continuing, new, and exiting students. These trends show the direction in which a program is moving. Changes in exit rates can reflect the effects of new policies and programs. For example, the overall change in exit rate for the NBTD spreadsheet shows a decrease of 5%. This is thought to be the result of changes in the first year student curriculum implemented in the previous two years. Changes in the entry rates of new students show where a program is gaining or losing admissions. The increase shown in new, year one NBTD students reflects the increase in NTID students from the Rubella epidemics of the early 1960's. Thus, it can be seen that these rates of change are immensely important for planning.

One additional spreadsheet function that would be useful to Student Flow modeling is a random number generator. When projecting student numbers, some program categories tend to naturally become so small that they run to zero. Once a category has gone to zero, no more students will be projected into it. However, in real life, some programs will run a small, if sporatic, enrollment that necessitates resource allocation. Although probabilities for enrollment can be hand-calculated, and cells changed to add students, automatic generation of random students at actual rates would be cleaner. This would also add an appealing dimension of realism to the projection process.

This thesis leads to several possibilites for future research topics. First, since enrollment projection is a complex process, it could be implemented as an expert system. The intricacies of flow data analysis, projection calculation, and model verification are handled best by a person with extensive experience in data processing and educational research. However, although most educational institutions need to be able to estimate future enrollments, for many of them the area responsable for the projections may not have such a skilled individual available to it. An expert system would increase the probability that a less skilled staff would produce valid projections. Also, there is no procedure in 20/20 to print out a meaningful listing of the calculations used in the spreadsheet. The only option available is to "export" the entire model from a range of cells to a file as though for transfer to another machine, and then print it. The file thus produced has one record for each cell exported, and the output is somewhat cryptic since it is really intended for 20/20 not people. The user must then wade through this file to find the formulas. A procedure to print formulas and their cell locations in a user-friendly format would be very helpful when verifying calculations or looking for problems. Finally, although spreadsheets are a very flexible tool, they are not satisfactorily self-documenting. A method for connecting comments or more extensive documentation with individual cells would be beneficial.

In conclusion, this project has used current computer technology to design and develop a mathematical simulation of Student Flow. DFDs were used to analyse the academic process and find the important student activities which must be simulated. Longitudinal student data was analyzed with the SPSS-X Statistical Package to assign students to flow groups and generate the yearly transition matrices for this simulation. Finally, the 20/20 Integrated Spreadsheet Modeling Program was used to store the transition matrices, and generate both the Rates of Flow Database and the actual enrollment projections. The

RESULTS AND SUMMATION

Student Flow simulation and enrollment projection developed here would be a useful tool for the enrollment management process of any academic department or institution.

APPENDIX A

GLOSSARY

Admission Yield = The ratio of the number of students who enroll in a program, college, or institution to the number of acceptances.

Base Term = A term from which Student Flow activity is measured. Continuing and exiting students are active in the program of interest during the base term. New students do not exist in the program in the base term. See "measure term".

Cross-Sectional Analysis = The study of a "narrow" problem, such as the results of one test or experiment, done by repeated measurement of the same variable(s) at different points in time.

Enrollment Management = The recruitment and retention strategies that are used by an educational institution to try to affect the type, quality, and number of students it has.

Longitudinal Analysis = A study of some process or activity that focuses on a static population over an extended period of time for the purpose of answering some general question(s). Longitudinal studies frequently point out areas in which cross sectional studies are needed.

Measure Term = A term at which Student Flow activity is measured. Continuing and new students are active in the program of interest in the measure term. Exiting students leave the program of interest in or before the measure term. See "Student Flow".

Student Flow = the proportions of students in various activity categories across successive base and measure terms.

Transition Matrix = a matrix of probabilities that a student in status X at time T will be in status Y at time T+1, where Y can equal X.

APPENDIX B

STUDENT HISTORY DATABASE DEFINITION

FIELD NAME	LOCATION	FORMAT	DESCRIPTION
SSNO	1-9	F9.0	student social security #
ENTRYQTR	10-12	F3.0	original entry quarter to RIT/NTID; standard RIT quarter codes
POSTSEC	13	A1	previous postsecondary education flag: ´Y´ = yes ´N´ = no
RACE	14	A1	race code: ´O´ = caucasian else = minority
CURRSTAT	16	A1	student´s current or exit status: ´R´,´C´,´B´,´F´ = active ´L´ = LOA ´G´ = graduated ´W´ = withdrawn ´I´ = inactive ´D´ = disciplinary suspension

FIELD NAME	LOCATION	FORMAT	DESCRIPTION
LASTDEGR	17	A1	most recent degree received: ´C´ = certificate ´D´ = diploma ´A´ = AAS or AOS ´B´ = bachelors ´M´ = masters
ADMTFLAG	18	F1.0	was student readmitted prior to Fall (801): 0 = no 1 = yes
CHNGFLAG	19	F1.0	did student change program prior to Fall (801): 0 = no 1 = yes
PRIORCR	20-22	F3.0	total number of credit hours accumulated towards graduation prior to Fall (801)
QTR1 to QTR21	23-85	21F3.0	quarter id´s for 801 through 851
STATUS1 to STATUS21	86-106	21A1	academic status for quarters 801 to 851; same codes as CURRSTAT excluding ´G´
PROG1 to PROG21	107-190	21A4	program code for quarters 801 to 851; standard RIT program codes
CREDHR1 to CREDHR21	191-253	21F3.0	credit hours earned towards graduation for quarters 801 to 851

APPENDIX C

SAMPLE PROJECTIONS

columns →

rows ↓

ENROLLMENT PROJECTION MODEL - PROJECTION FOR: NTID DATA PROCESSING 8/6/86

NBID 5 YEAR ENROLLMENT HISTORY DATA

Row	YEAR IN PROGRAM	ENTRY STATUS	YEAR 0 ALL ACTIVE STUDENTS N	YEAR 0 ALL ACTIVE STUDENTS .P	YEAR 0-1 EXITING STUDENTS N	YEAR 0-1 EXITING STUDENTS .P	YEAR 1 CONTINUING STUDENTS N	YEAR 1 CONTINUING STUDENTS .P	YEAR 1 NEW STUDENTS N	YEAR 1 NEW STUDENTS .P
12	1	NEW	50	0.379	20	0.400	2	0.028	35	0.530
13		EXT. TR	5	0.038	0	0.000	1	0.014	17	0.258
14		INT. TR	5	0.038	3	0.600	0	0.000	10	0.152
15	2	NEW	31	0.235	12	0.387	29	0.408	0	0.000
...		EXT. TR	0	0.000	0	0.000	4	0.056	0	0.000
		INT. TR	12	0.091	4	0.333	3	0.042	4	0.061
	3	NEW	15	0.114	10	0.667	17	0.239	0	0.000
		EXT. TR	6	0.045	5	0.833	0	0.000	0	0.000
		INT. TR	1	0.008	1	1.000	6	0.085	0	0.000
	4	NEW	6	0.045	5	0.833	6	0.085	0	0.000
		EXT. TR	0	0.000	0	0.000	1	0.014	0	0.000
		INT. TR	0	0.000	0	0.000	1	0.014	0	0.000
	5	NEW	1	0.008	1	1.000	1	0.014	0	0.000
		EXT. TR	0	0.000	0	0.000	0	0.000	0	0.000
		INT. TR	0	0.000	0	0.000	0	0.000	0	0.000
	6	NEW	0	0.000	0	0.000	0	0.000	0	0.000
		EXT. TR	0	0.000	0	0.000	0	0.000	0	0.000
		INT. TR	0	0.000	0	0.000	0	0.000	0	0.000
	7	NEW	0	0.000	0	0.000	0	0.000	0	0.000
		EXT. TR	0	0.000	0	0.000	0	0.000	0	0.000
32		INT. TR	0	0.000	0	0.000	0	0.000	0	0.000
34		TOTAL STUDENTS	132 (SUM=	1.000)	61	0.462	71 (SUM=	0.538 1.000)	66 (SUM=	0.482 1.000)

Row	ASSUMPTIONS:	YEAR +1	YEAR +2	YEAR +3	YEAR +4	YEAR +5	WEIGHTS:	SUM	YEARS 0-1	YEARS 1-2	YEARS 2-3	YEARS 3-4	YEARS 4-5
40	NEW STUDENTS	70	60	0	0	0	EXITING	5.0	1.0	1.0	1.0	1.0	1.0
41							NEW	5.0	1.0	1.0	1.0	1.0	1.0
42							CON'T.	5.0	1.0	1.0	1.0	1.0	1.0

Enrollment History Data

Columns →　　*rows ↓*

Rows	YEAR 1-2		YEAR 2				YEAR 2-3		YEAR 3			
	EXITING STUDENTS		CONTINUING STUDENTS		NEW STUDENTS		EXITING STUDENTS		CONTINUING STUDENTS		NEW STUDENTS	
Columns	15	16	18	19	20	21	23	24	26	27	28	29
	N	.P	N	.P	N	.P	N	.P	N	.P	N	.P
12	15	0.405	2	0.026	46	0.622	22	0.458	5	0.056	79	0.731
13	5	0.278	4	0.052	4	0.054	1	0.125	0	0.000	10	0.093
14	2	0.200	0	0.000	16	0.216	4	0.250	2	0.022	12	0.111
15	11	0.379	19	0.247	1	0.014	6	0.300	20	0.225	0	0.000
...	1	0.250	9	0.117	0	0.000	3	0.333	5	0.056	0	0.000
	6	0.857	10	0.130	4	0.054	4	0.286	14	0.157	5	0.046
	9	0.529	17	0.221	0	0.000	6	0.353	15	0.169	0	0.000
	0	0.000	3	0.039	0	0.000	1	0.333	7	0.079	0	0.000
	5	0.833	1	0.013	1	0.014	1	0.500	7	0.079	1	0.009
	3	0.500	7	0.091	0	0.000	3	0.429	9	0.101	0	0.000
	1	1.000	0	0.000	0	0.000	0	0.000	1	0.011	0	0.000
	0	0.000	2	0.026	2	0.027	1	0.250	1	0.011	1	0.009
	1	1.000	3	0.039	0	0.000	0	0.000	2	0.022	0	0.000
	0	0.000	0	0.000	0	0.000	0	0.000	0	0.000	0	0.000
	0	0.000	0	0.000	0	0.000	0	0.000	0	0.000	0	0.000
	0	0.000	0	0.000	0	0.000	0	0.000	1	0.011	0	0.000
	0	0.000	0	0.000	0	0.000	0	0.000	0	0.000	0	0.000
	0	0.000	0	0.000	0	0.000	0	0.000	0	0.000	0	0.000
	0	0.000	0	0.000	0	0.000	0	0.000	0	0.000	0	0.000
	0	0.000	0	0.000	0	0.000	0	0.000	0	0.000	0	0.000
32	0	0.000	0	0.000	0	0.000	0	0.000	0	0.000	0	0.000
34	59	0.431	77 (SUM=	0.562 1.000)	74 (SUM=	0.490 1.000)	52	0.344	89 (SUM=	0.589 1.000)	108 (SUM=	0.548 1.000

History Data (con 4.)

columns →

rows ↓

	31	32	34	35	36	37	39	40	42	43	44	45
	YEAR 3-4		YEAR 4				YEAR 4-5		YEAR 5			
	EXITING STUDENTS		CONTINUING STUDENTS		NEW STUDENTS		EXITING STUDENTS		CONTINUING STUDENTS		NEW STUDENTS	
	N	.P	N	.P	N	.P	N	.P	N	.P	N	.P
12	18	0.214	5	0.039	65	0.596	15	0.214	8	0.047	40	0.615
13	2	0.200	0	0.000	7	0.064	1	0.143	0	0.000	13	0.200
14	4	0.286	2	0.016	26	0.239	6	0.214	2	0.012	4	0.062
15	0	0.000	59	0.465	3	0.028	6	0.097	53	0.310	0	0.000
.	2	0.400	7	0.055	0	0.000	1	0.143	6	0.035	0	0.000
.	8	0.421	8	0.063	5	0.046	4	0.308	19	0.111	6	0.092
.	8	0.533	18	0.142	1	0.009	12	0.632	48	0.281	0	0.000
	3	0.429	4	0.031	0	0.000	2	0.500	5	0.029	0	0.000
	3	0.375	8	0.063	1	0.009	7	0.778	9	0.053	1	0.015
	4	0.444	7	0.055	0	0.000	4	0.571	6	0.035	0	0.000
	0	0.000	3	0.024	0	0.000	1	0.333	3	0.018	0	0.000
	2	1.000	3	0.024	1	0.009	0	0.000	3	0.018	1	0.015
	0	0.000	2	0.016	0	0.000	2	1.000	5	0.029	0	0.000
	0	0.000	0	0.000	0	0.000	0	0.000	0	0.000	0	0.000
	0	0.000	0	0.000	0	0.000	0	0.000	4	0.023	0	0.000
	0	0.000	0	0.000	0	0.000	0	0.000	0	0.000	0	0.000
	0	0.000	0	0.000	0	0.000	0	0.000	0	0.000	0	0.000
	0	0.000	0	0.000	0	0.000	0	0.000	0	0.000	0	0.000
	0	0.000	1	0.008	0	0.000	1	1.000	0	0.000	0	0.000
	0	0.000	0	0.000	0	0.000	0	0.000	0	0.000	0	0.000
32	0	0.000	0	0.000	0	0.000	0	0.000	0	0.000	0	0.000
34	54	0.274	127 (SUM=	0.645 1.000)	109 (SUM=	0.462 1.000)	62	0.263	171 (SUM=	0.725 1.000)	65 (SUM=	0.275 1.000)

History Data (con't.)

columns →

rows ↓

CALCULATIONS:

Row			3	5	7	9	11	13
	YEAR IN PROGRAM	ENTRY STATUS	AVERAGE EXITING YEARS 1-5	AVERAGE NEW YEARS 1-5	AVERAGE CHANGE FOR EXITING YEARS 1-5	AVERAGE CHANGE FOR NEW YEARS 1-5	AVERAGE CONTINUING YEARS 1-5	AVERAGE CHANGE CONTINUING YEARS 1-5
			.P	.P	.P	.P	.P	.P
50	1	NEW	0.338	0.619	-0.0464	0.0213	0.039	0.0047
51		EXT. TR	0.149	0.134	0.0357	-0.0144	0.013	-0.0035
52		INT. TR	0.310	0.156	-0.0964	-0.0225	0.010	0.0029
53	2	NEW	0.233	0.008	-0.0726	0.0000	0.331	-0.0246
.		EXT. TR	0.225	0.000	0.0357	0.0000	0.064	-0.0053
.		INT. TR	0.441	0.060	-0.0064	0.0079	0.101	0.0172
.	3	NEW	0.543	0.002	-0.0088	0.0000	0.210	0.0103
		EXT. TR	0.419	0.000	-0.0833	0.0000	0.036	0.0073
		INT. TR	0.697	0.009	-0.0556	0.0038	0.058	-0.0080
	4	NEW	0.556	0.000	-0.0655	0.0000	0.073	-0.0124
		EXT. TR	0.267	0.000	0.0833	0.0000	0.013	0.0009
		INT. TR	0.250	0.012	0.0000	0.0038	0.018	0.0009
	5	NEW	0.600	0.000	0.0000	0.0000	0.024	0.0038
		EXT. TR	0.000	0.000	0.0000	0.0000	0.000	0.0000
		INT. TR	0.000	0.000	0.0000	0.0000	0.005	0.0058
	6	NEW	0.000	0.000	0.0000	0.0000	0.002	0.0000
		EXT. TR	0.000	0.000	0.0000	0.0000	0.000	0.0000
		INT. TR	0.000	0.000	0.0000	0.0000	0.000	0.0000
	7	NEW	0.200	0.000	0.2500	0.0000	0.002	0.0000
		EXT. TR	0.000	0.000	0.0000	0.0000	0.000	0.0000
70		INT. TR	0.000	0.000	0.0000	0.0000	0.000	0.0000
72	OVERALL AVERAGE		0.355	0.451 (SUM = 1.0000)	-0.0499	-0.0516	0.612 (SUM = 1.0000)	0.0467

Rates of Flow Database

Columns →

Rows ↓

Row	YEAR IN PROGRAM	ENTRY STATUS	YEAR +1 CONTINUING STUDENTS (N) (3)	YEAR +1 NEW STUDENTS (N) (5)	YEAR +1 TOTAL STUDENTS (N) (7)	YEAR +2 CONTINUING STUDENTS (N) (9)	YEAR +2 NEW STUDENTS (N) (11)	YEAR +2 TOTAL STUDENTS (N) (13)
	PROJECTIONS:	2 YEARS FORWARD						
81	1	NEW	7	45	52	7	40	47
82		EXT. TR	1	8	10	1	6	7
83		INT. TR	2	9	11	2	7	9
84	2	NEW	47	1	47	40	0	41
.		EXT. TR	9	0	9	8	0	8
.		INT. TR	18	5	23	19	5	24
.	3	NEW	34	0	34	33	0	33
		EXT. TR	7	0	7	7	0	7
		INT. TR	8	1	9	6	1	7
	4	NEW	9	0	9	7	0	7
		EXT. TR	2	0	2	2	0	2
		INT. TR	3	1	4	3	1	4
	5	NEW	4	0	4	5	0	5
		EXT. TR	0	0	0	0	0	0
		INT. TR	2	0	2	2	0	2
	6	NEW	0	0	0	0	0	0
		EXT. TR	0	0	0	0	0	0
		INT. TR	0	0	0	0	0	0
	7	NEW	0	0	0	0	0	0
		EXT. TR	0	0	0	0	0	0
101		INT. TR	0	0	0	0	0	0
103	TOTAL STUDENTS		152	70	222	143	60	203

Enrollment Projections for Two Years Forward

APPENDIX D

PROJECTION FORMULAS

ENROLLMENT PROJECTION FORMULAS

(KEY: cell location = [column,row])

20/20 SPREADSHEET LOCATION	FORMULA	DESCRIPTION
ENROLLMENT HISTORY: (Annual Transition Matrices)		
column 4, rows 12..32	[3,x] / [3,34]; x=row 12..32	proportions of active students year 0 by entry status and year in program
column 8, rows 12..32	if([3,x] > 0) then [7,x] / [3,x] else 0; x=rows 12..32	proportions of students exiting before year 1 by year 0 category
column 11, rows 12..32	[10,x] / [10,34]; x=12..32	proportion of continuing students in year 1 by category
cell [11,34]	[10,34] / [3,34]	overall proportion of students continuing from year 0
column 13, rows 12..32	[12,x] / [12,34]; x=12..32	proprotions of new students by entry status and year in program year 1
cell [13,34]	[12,34] / ([10,34] + [12,34])	proportion of the class year 1 who are new
column 16, rows 12..32	if([10,x] + [12,x] = 0) then 0 else [15,x] / ([10,x] + 12,x]); x=row 12..32	proportions of year 1 students who exited before year 2 by category
cell [16,34]	[15,34] / ([10,34] + [12,34])	overall exit rate from year 1
column 19, rows 12..32	* same as column 11 for year 2 *	

20/20 SPREADSHEET LOCATION	FORMULA	DESCRIPTION

ENROLLMENT HISTORY (con´t.):

cell [19,34]	* same as cell [11,34] for year 2 *	
column 21, rows 12..32	* same as column 13 for year 2 *	
cell [21,34]	* same as cell [13,34] for year 2 *	
column 24, rows 12..32	* same as column 16 for years 2-3 *	
cell [24,34]	* same as cell [16,34] for years 2-3 *	
column 27, rows 12..32	* same as column 11 for year 3 *	
cell [27,34]	* same as cell [11,34] for year 3 *	
column 29, rows 12..32	* same as column 13 for year 3 *	
cell [29,34]	* same as cell [13,34] for year 3 *	
column 32, rows 12..32	* same as column 16 for years 3-4 *	
cell [32,34]	* same as cell [16,34] for years 3-4 *	
column 35, rows 12..32	* same as column 11 for year 4 *	
cell [35,34]	* same as cell [11,34] for year 4 *	
column 37, rows 12..32	* same as column 13 for year 4 *	
cell [37,34]	* same as cell [13,34] for year 4 *	
column 40, rows 12..32	* same as column 16 for years 4-5 *	
cell [40,34]	* same as cell [16,34] for years 4-5 *	
column 43, rows 12..32	* same as column 11 for year 5 *	

20/20 SPREADSHEET LOCATION	FORMULA	DESCRIPTION
ENROLLMENT HISTORY (con't.):		
cell [43,34]	* same as cell [11,34] for year 5 *	
column 45, rows 12..32	* same as column 13 for year 5 *	
cell [45,34]	* same as cell [13,34] for year 5 *	
ASSUMPTIONS:		
cell [2,40]	constant	# new students projected for year +1
cell [3,40]	constant	# new students projected for year +2
cells [4,40] to [6,40]	* same as [2,40] for years +3 to +5 *	
WEIGHTS:		
cell [9.40]	sum([10..14,40])	sum of all weighting factors
cell [10.40]	constant	weighting factor for years 0-1 exiting
cells [11,40] to [14.40]	* same as [10,40] for years 1-2 to 4-5 *	
cells [9,41] to [14,41]	* same as [9.40] to [14.40] for new students *	
cells [9,42] to [14.42]	* same as [9,40] to [14,40] for continuing students *	

20/20 SPREADSHEET LOCATION	FORMULA	DESCRIPTION
CALCULATIONS: (Rates of Flow Database)		
column 3, rows 50..70	sum([8,x]*[10,40],[16,x]*[11,40], [24,x]*[12,40],[32,x]*[13,40], [40,x]*[14,40])/[9,40]; x=row 12..32	average exiting by category across 5 years
cell [3,72]	sum([8,34]*[10,40],[16,34]*[11,40], [24,34]*[12,40],[32,34]*[13,40], [40,43]*[14,40])/[9,40]	overall average exiting
column 5, rows 50..70	* same as 3,50..70 for new students *	
cell [5,72]	* same as [3,72] for new students *	
column 7, rows 50..70	sum(([16,x]-[8,x])*[11,40], ([24,x]-[16,x])*[12,40], ([32,x]-[24,x])*[13,40], ([40,x]-[32,x])*[14,40]) / ([9,40]-[10,40]); x=row 12..32	average rate of change of exiting across 5 years
cell [7,72]	sum(([16,34]-[8,34])*[11,40], ([24,34]-[16,34])*[12,40], ([32,34]-[24,34])*[13,40], ([40,34]-[32,34])*[14,40]) / ([9,40]-[10,40])	overall rate of change of exiting across 5 years
column 9, rows 50..70	* same as column 7 for new students *	
cell [9,72]	* same as cell [7,72] for new students *	
column 11, rows 50..70	* same as column 3 for continuing students *	
cell [11,72]	* same as cell [3,72] for continuing students *	
column 13, rows 50..70	* same as column 7 for continuing students *	
cell [13,72]	* same as cell [7,72] for continuing students *	

20/20 SPREADSHEET LOCATION	FORMULA	DESCRIPTION
PROJECTIONS:		
column 3, rows 81..101	if([11,x] + [13,x] > 0) then (([42,34] + [44,34]) * (1.0-[3,72])) * ([11,x] + [13,x]) else 0; x=row 50..70	projected # continuing students by category year +1
cell [3,103]	sum([3,81..101])	total continuing students year +1
column 5, rows 81..101	if([5,x] + [9,x] > 0) then [2,40] * ([5,x] + [9,x]) else 0; x=row 50..70	projected # new students by category year +1
cell [5,105]	sum([5,81..101])	total new students year +1
column 7, rows 81..101	[3,x] + [5,x]; x=row 81,101	total projected students by category year +1
cell [7,103]	sum([7,81..101])	total projected students year +1
column 9, rows 81..101	if([11,x] + (2 * [13,x]) > 0) then (([7,103] * (1.0 - [3,72])) * ([11,x] + (2 * [13,x])) else 0; x=50..70	projected # continuing students by category year +2
cell [9,103]	sum([9,81..101])	total continuing students year +2
column 11, rows 81..101	if([5,x] + (2 * [9,x]) > 0) then [3,40] * ([5,x] + (2 * [9,x])) else 0; x=row 50..70	projected # new students by category year +2
cell [11,103]	sum([11,81..101])	total new students year +2
column 13, rows 81..101	[9,x] + [11,x]; x=row 81..101	total projected students by category year +2
cell [13,103]	sum([13,81..101])	total projected students year +2

APPENDIX E

BIBLIOGRAPHY

1 Atre, S., "Strategic Planning in Data Base Technology", presentation at the National Database and 4th. Generation Language Symposium, Teaneck, New Jersey, June, 1986.

2 Balderston, F. E., "The Design and Uses of Information Systems", in Bowen, H. R. (ed.), Evaluating Institutions for Accountability, Josse-Bass Inc., San Francisco, 1974, pp. 47-66.

3 Buffa, E. and Dyer, J., Management Science/Operations Research, John Wiley & Sons, Santa Barbara, 1977, pp. 143-147.

4 DeMarco, T., Structured Analysis and System Specification, Yourdon Press, New York, 1978.

5 DeMarco, T., Concise Notes on Software Engineering, Yourdon Press, New York, 1979.

6 EXCELERATOR Product Description, InTech, 5 Cambridge Center, Cambridge, MA 02142.

7 Fishbein, M. and Ajzen, I., Belief, Attitude, Intention and Behavior: An Instrduction to Theory and Research, Addison-Wesley, Mass., 1975.

8 Gane, C. and Sarson, T., Structured Systems Analysis: Tools and Techniques, IST, Inc., New York, 1977.

9 Giffin, W., Introduction to Operations Engineering, Richard D. Irwin, Inc., Illinois, 1971, pp. 591-593.

10 Glover, R. H., "Designing a Decision-Support System for Enrollment Management", Research in Higher Education, Vol. 24, 1986, pp. 15-34.

11 Hollander, T. E., "Planning for Changing Demographic Trends in Public and Private Institutions", (ERIC ED 109 961), 1974.

12 Lyell E. and Toole, P., "Student Flow Modelling and Enrollment Forecasting", Planning for Higher Education, Vol. 3, No. 6: 2/6, December 1974.

13 Martin, J. and McClure, C., Diagramming Techniques for Analysts and Programmers, Prentice-Hall, New Jersey, 1985, pp. 93-108

14 Moder, J. and Elmaghraby, S., Handbook of Operations Research, Foundations and Fundamentals, Van Nostrand Reinhold Co., New York, 1978, pp. 566-568.

15 Pascarella, E., Studying Student Attrition, Jossey-Bass Inc., San Francisco, 1982.

16 STRADIS/DRAW Product Description, MCAUTO, McDonnell Douglas Automation Company, Box 516, Dept. K277, Saint Louis, MO 63166.

17 Tinto, V., "Dropout from Higher Education: A Theoretical Synthesis of Recent Research", Review of Educational Research, 1975, 45, pp. 89-125.

18 Walter, G., "Length of Stay as a Function in Enrollment Size", Annual Conference North East Association for Institutional Research, Hartford, Conneticut, October, 1985.

19 Wasik, J. L., "The Development of a Mathematical Model to Project Enrollments in a Community College System", (ERIC ED 047 674), 1971.

20 Weinberg, V., Structured Analysis, Yourdon Press, New York, 1979, pp. 55-82.

21 20/20 Integrated Spreadsheet Modeling Program, Access Technology Inc., 6 Pleasant St., South Natick, Mass., 01760, pp. v-vi.