DOCUMENT RESUME

ED 062 729 24 EA 004 344

AUTHOR Denham, Carolyn H.
TITLE Probabilistic School Enrollment Predictions Using Monte Carlo Computer Simulation. Final Report.
INSTITUTION Boston Coll., Chestnut Hill, Mass.
SPONS AGENCY National Center for Educational Research and Development (DHEW/OE), Washington, D.C.
BUREAU NO BR-9-A-024
PUB DATE May 71
GRANT OEG-1-9-090024-0107(010)
NOTE 171p.

EDRS PRICE MF-$0.65 HC-$6.58
DESCRIPTORS *Computer Oriented Programs; Enrollment Influences; *Enrollment Projections; Enrollment Trends; Literature Reviews; Prediction; Predictive Validity; Probability; Reliability; *School Demography; *Simulation; *Student Enrollment; Validity

ABSTRACT

A major difficulty in predictions of school enrollments is the failure of the forecaster to express adequately his degree of certainty in his estimates. To alleviate this problem, a method was developed by which a forecaster could prepare probability distributions of enrollment predictions. A basic method of enrollment prediction was chosen and modified to accommodate probabilistic input and output. The method required separate estimates for such variables as migration, retention, and transfer; and it was modified to require three estimates (high, low, and most likely) for each variable. A Monte Carlo computer simulation program was written to combine these various estimates into probability distributions of enrollment prediction. (Appendix A, pages 150-161, may reproduce poorly.) (Author/RA)

ED 062 729

FINAL REPORT
Project No.9-A-024
Grant No. OEG-1-9-090024-0107 (010)

PROBABILISTIC SCHOOL ENROLLMENT PREDICTIONS USING MONTE CARLO COMPUTER SIMULATION

Carolyn H. Denham
Boston College
Chestnut Hill, Massachusetts

May 1971

The research reported herein was performed pursuant to a grant with the Office of Education, U.S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

U. S. Department of Health, Education, and Welfare

Office of Education
Bureau of Research

EA 004 344

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

ACKNOWLEDGMENTS

John A. Schmitt of Boston College served as director of the project. I wish to thank him for suggesting the topic to me and for his guidance and support during the research. Other members of the Boston College faculty who were most helpful were Mary D. Griffin, Oscar J. Alers, Vincent C. Nuccio, and Ronald L. Nuttall.

PROBABILISTIC SCHOOL ENROLLMENT PREDICTIONS USING MONTE CARLO COMPUTER SIMULATION

Abstract

A major difficulty in predictions of school enrollments is the failure of the forecaster to express adequately his degree of certainty in his estimates. To alleviate this problem a method was developed by which a forecaster could prepare probability distributions of enrollment predictions.

A basic method of enrollment prediction was chosen and modified to accommodate probabilistic input and output; the method required separate estimates for variables such as migration, retention, and transfers. The method was modified to require three estimates for each variable: a high, a low, and a most likely estimate, with the high and the low estimates representing the 98 percent confidence interval. A Monte Carlo computer simulation program was written to combine these various estimates into probability distributions of enrollment prediction. Significance tests were performed to investigate predictive validity, reliability, and concurrent validity. Results indicated adequate reliability and validity, contingent upon additional tests, in all but

one of the tests, a test of concurrent validity. The results of this test suggest the need for re-examining the assumptions about the distributions of the probabilistic input.

The computer programs are ready for use although the user is encouraged to conduct additional tests of validity and reliability and to suggest improvements in the model. The model is unique among enrollment prediction methods in that it requires the user to examine the various parts of the system, to estimate probabilities for each of these parts, and to use the probabilities to determine probabilistic information about the operation of the system as a whole.

TABLE OF CONTENTS

		Page
Chapter I.	Introduction	1
Chapter II.	Review of the Literature	11
Chapter III.	Design of the Study	50
Chapter IV.	Results of the Study	97
Chapter V.	Summary and Conclusions	133
References		141
Appendixes		149
	a. Program MAIN	150
	b. Subroutine OUTPUT	160
	c. Instructions to the User	162

Tables and Figures

Table 1.1	Output Format	8
Figure 2.1	Multivariable model used in the present study	38
Figure 3.1	Multivariable model for grade 8 used by the Center for Field Studies	54
Figure 3.2	Multivariable model for grade 8 used in the present study	55
Figure 3.3	A beta distribution	64
Figure 3.4	Abridged flow chart for Program MAIN	78
Figure 3.5	Flow Chart for Program MAIN	79
Figure 3.6	Flow Chart for Subroutine OUTPUT	87

Page

Table 4.1 Sample output.....98

Table 4.2 Sample output, sex option.....99

Table 4.3 The comparison of the actual Brockton enrollments with those predicted by the multivariables technique and those projected by the percentage of survival technique.....109

Table 4.4 The 0.50 probability figures calculated by simulation with symmetrical data using different seeds for the random number generator.....112

Table 4.5 The comparison of the non-simulation multivariable predictions with the 0.50 probability figures for the symmetrical and skewed data.....117

Table 4.6 Means and corresponding percentages computed for data used in the tests of the four null hypotheses.....122

Table 5.1 Number of times the actual enrollments exceeded the .05 and .95 probability points of prediction....137

Chapter I

Introduction

Systems for predicting school enrollments constitute important decision making tools for school management. The decisions are important, not only because millions of dollars are involved, but because the decisions can greatly affect the quality of education school children receive. An underestimate of enrollment may result in crowded classrooms or double sessions. Lawrence Derthick (1957) testified that double sessions cause children to lose up to two months of schooling a year with a corresponding drop in achievement. An overestimate may result in a loss of money; obvious inefficiency may make the community less willing to support future building plans.

The primary purpose of most school enrollment predictions is to determine the extent, urgency, and immediacy of plant capacity needs. In planning new school buildings, decisions have to be made with respect to space requirements: short-range decisions on classroom space requirements and long-range decisions on special facilities such as cafeterias,

auditoriums, and playgrounds. Predictions of probable extent and timing of peak enrollment or changes in enrollment distribution by grade indicate the need for flexibility in the school plant. Cost estimates must be made and the exact time and place for the erection of buildings must be chosen. Related questions are the amount of money required for the school buildings, district boundaries, grade level organization, and probable future functions of the school plant (Larson & Strevell, 1952:65-66).

In addition to space requirements, short-range enrollment predictions are useful in making plans for obtaining specialists such as counselors or nurses, providing services for exceptional students such as the handicapped, insuring an adequate teacher-pupil ratio, and purchasing equipment such as teaching machines and language laboratory terminals.

Besides providing information for planning a school building, the predictions can be useful in facilities planning for the school system as a whole. They provide a basis for predicting the amount of bonded indebtedness that will have to be incurred. Needed school sites can be anticipated; school sites and boundaries can be chosen for optimum distribution of the population and, if the predictions include the necessary

information, distribution of racial and ethnic groups. Predictions can be used to indicate when use of substandard buildings may be discontinued.

There is agreement among the forecasters and users of enrollment predictions that dependable predictions have a substantial influence on the future direction and quality of educational programs. Many also realize that predictions of school enrollments are a hazardous task; dependable results are difficult to guarantee.

Two of the problems making the task difficult are the unpredictability of the phenomena and the inaccuracy of the prediction methods used. It is a major thesis of this study that another part of the difficulty is the lack of understanding of the prediction results by the user. Often the user of the prediction figures is not the originator of these figures. Although school boards and administrators are most frequently the users of the predictions, the predictions are often made by citizens' committees (e.g., Population and Housing Committee, 1966; Citizens Advisory Committee on School Needs, 1960), by the research personnel of the school system, or by consultants (e.g., Marshall, 1968; Arthur D. Little, Inc., 1966). The user is often unaware

of the assumptions underlying the predictions. More often he is unaware of the degree of confidence the forecaster has in his predictions. The forecaster may be aware that the actual enrollment is unlikely to be exactly as predicted, but it is not uncommon for the user to place too much confidence in single figure predictions. The method of producing enrollment predictions developed in the present study was, to a large extent, designed to overcome the lack of communication between the forecaster and the user.

Enrollment predictions may be made and presented as a single figure, as a high and a low figure, or as three or more figures based on different assumptions about the future enrollment. In the latter two cases, the figures may or may not be accompanied by an indication of the probability that the actual enrollment figure will fall on or between certain figures. A consideration in choosing one of these alternatives for the present study was the potential for communication to the user of the extent of forecaster certainty. A single figure may give the user unwarranted confidence in the dependability of the forecast (Stanbery, 1952:10). One solution to the communication problem was used by Marshall (1968). He explained in words

the uncertainties and the most probable departures from the predictions. He stated that "any future change in zoning, municipal services, residential development, or parochial schooling may send enrollment off in a direction higher or lower than that projected here [p. A-1]." He also stated that his projections for elementary school enrollment for 1975 and 1980 are "probably conservative [p. A-9]." However, when the consultant is hired to produce prediction figures, such disclaimers do not have the same impact as do the prediction figures themselves. Furthermore, if the user is going to employ this added information in his planning, he must translate the words into approximate numbers or ranges of numbers. Assuming that it is the forecaster who has the information and expertise for making these numerical estimates, it would be desirable for him to present his prediction figures in such a way as to communicate his extent of certainty in his predictions or his estimates of probabilities of various enrollments or ranges of enrollment.

Stanbery (1952:10-12) discussed two figure presentations (a high and a low figure, representing a probable range) and multiple figure presentations as methods of communicating

information to the user. He recommended use of two predictions, one high and one low, rather than multiple predictions: he reasoned that a presentation of multiple predictions is unwieldy and that it is difficult for the user to know which of the predictions is most likely to be realized. Multiple predictions are usually statistical computations of enrollments that would result under various assumptions. However, a probability distribution of predictions would be a way in which the forecaster could communicate to the user his judgment of probabilities of various outcomes in a manner which is not "unwieldy." Stanbery (1952) wrote:

> The factors and conditions affecting population change...are so numerous and their effects are so varied that it would be impractical to assign proper weights to each of them and to compute mathematically the probabilities of realizing any one figure between the maximum and minimum projections [p. 45].

In the present study a method was developed for the purpose of computing probabilities for enrollment in a way which would not be "impractical." The method developed is essentially a modification of an existing method of enrollment prediction to accommodate probabilistic input and to produce probabilistic output. In order to modify the basic prediction method, computer simulation was used. The simulation produces

probability distributions of enrollments expressed in the format given in Table 1.1 (p. 8).

The model which was used as the basis of the simulation is referred to as the "multivariable" method (Johnson, 1965) and has been used in various prediction studies (e.g., Center for Field Studies, 1964; Center for Field Studies, 1956).

The model requires that separate predictions be made for major factors affecting school enrollment: births, migrations, retentions, transfers to and from nonpublic schools, school dropouts, and deaths. Enrollment predictions for a certain year and grade are computed by adjusting the enrollment figures for the previous year and grade by adding or subtracting, as appropriate, the predicted values of each of the factors during the previous year to obtain an estimate for the year and grade under construction. This model was modified for use in the simulation by requiring the forecaster to estimate high, most likely, and low figures for each of the variables, with the high and low estimates representing the limits of the 98 percent confidence interval. For example, the high estimate of grade 8 migration between October 1, 1970 and October 1, 1971, may be 15; the most likely estimate, 10; and the low estimate, 5. These three numbers are assumed to describe a probability distribution.

TABLE 1.1

PROBABILITY THAT TOTAL ENROLLMENT IN GRADE 1 IN 1974 WILL BE LESS THAN THE SPECIFIED PREDICTED ENROLLMENT

PROBABILITY	PREDICTED ENROLLMENT
.05	2023
.10	2100
.20	2171
.30	2244
.40	2316
.50	2384
.60	2429
.70	2513
.80	2549
.90	2653
.95	2735

PROBABILITY THAT TOTAL ENROLLMENT IN GRADE 1 IN 1974 WILL BE GREATER THAN THE SPECIFIED PREDICTED ENROLLMENT

PROBABILITY	PREDICTED ENROLLMENT
.05	2735
.10	2653
.20	2549
.30	2513
.40	2429
.50	2384
.60	2316
.70	2244
.80	2171
.90	2100
.95	2023

The computer simulation is used as a means of combining the distributions of the various factors into probability distributions of predicted enrollments.

The multivariable prediction method was chosen because it allows the forecaster freedom to deviate from mere projection of past data and allows him to make probability statements about individual variables such as migration. The literature on school enrollment prediction has very few good validity studies, but there are reasonable arguments for preferring the multivariable method to other recognized prediction methods (Whitla, 1952).

The data used in the study consisted of enrollment predictions made in 1964 for the City of Brockton, Massachusetts. This use of previously made predictions allowed for a test of predictive validity; the accuracy of the prediction using the multivariable method was compared to that using the percentage of survival method, probably the most popular prediction method. A second significance test was used to test the reliability of the simulation output when the starting value of the random number generator is varied; the hypothesis was that there is no difference between output of the simulation using two different starting values for the random number generator.

Concurrent validity was investigated by comparing the non-simulation multivariable predictions to the distributions produced by the simulation using two different types of input data.

In summary, the study developed a method for producing probability distributions of school enrollment predictions; a basic method for single figure predictions was chosen, and Monte Carlo computer simulation programs were written to produce multiple predictions in the form of distributions. Significance tests were performed to investigate predictive validity of the prediction model and reliability and concurrent validity of the simulation output.

Chapter II

Review of the Literature

We do not have methods which can predict precisely the enrollment in a certain grade and year. Our methods are not that refined; our prescience as to the future events involved is not that accurate. For a prediction study to be complete it should include an indication of the forecaster's certainty of the predictions. Whether or not the user is also the forecaster, the user of the enrollment predictions should take account of the assumptions and uncertainties in making his plans. It is the contention here that estimation and communication of this information is more precise, forceful, and useful if presented in numbers or graphs rather than in words.

Stanbery (1952:10-12) discussed the relative advantages and disadvantages of single, double, and multiple figure predictions for these purposes. Among the disadvantages he cited for single figure predictions are the unwarranted confidence in the accuracy of the prediction it gives the user, its failure to give the user any indication of the extent to which it might be in error, and the fact that it may be based on only one set of assumptions. He stated that double figure

predictions, one high and one low figure, meet these objections and allow the user to compare the assumptions on which the high and low predictions are based and to make his own judgment about probabilities of the low figure, the high figure, or some intermediate figure being realized. However, Stanbery also realized that it might indeed be a disadvantage to make it necessary for the user to exercise his judgment in this way. Advantages cited for the multiple figure predictions were much the same as those cited for double figure predictions. Multiple predictions are usually presented as statistical computations of the populations that would result under various assumptions about such variables as birth rate, death rate, and migration. He cited as a disadvantage the fact that little guidance is usually given to the user about which of the assumptions are more likely to be correct, and many users get the impression that one of the intermediate figures is more likely to be realized than one of the higher or lower figures, an assumption which is not necessarily correct. Also, Stanbery described multiple predictions as impractical and unwieldy. Consideration of the foregoing advantages and disadvantages of the various methods led Stanbery to recommend the use of double figure predictions.

A type of prediction that Stanbery did not consider is a probability distribution of predictions. Through the addition of probabilities to the predictions the forecaster indicates to the user which sets of assumptions he feels are most likely to be correct. Of course, the forecaster can also report the basic assumptions on which the probabilities are based, giving the user an opportunity to review and possibly modify the judgments of the forecaster. The probability distribution of predictions, however, takes the burden of choosing among various assumptions from the user and gives it to the forecaster, who presumably has access to more information. The present study is an investigation of a method for producing multiple figure predictions which are not "impractical" and"unwieldy" by producing probability distributions of enrollment predictions using computer simulation.

It is the present investigator's contention that there is a need for a satisfactory method of preparing predictions in the form of probability distributions although the application of probabilities to predictions has been reported in demographic literature. Stanbery (1952:12) recommended that high and low figures be chosen to produce a range which can be expected with the probability of .50 or greater to contain

the actual population. Griffin and Schmitt (1966) proposed the use of Monte Carlo computer simulation to produce probabilistic output. Peters (1969) is developing a method of calculating confidence intervals for school enrollment projections, with the intervals being the 90 or 95 percent confidence intervals. This is perhaps the most extensive attempt to produce probabilistic output. However, Peters' method is based on the percentage of survival method, a method of forecasting which the present investigator considers unsatisfactory in that it fails to allow for changes in trends.

In the present study the multivariable prediction method was modified to accommodate probabilistic input and output. Since the purpose of the study was not to develop new demographic forecasting techniques, but to develop a way to account for probabilities in predictions, a previously existing prediction method was selected and modified.

A review of the literature showed no single taxonomy of enrollment prediction method to be adequate as a basis for discussing alternative prediction methods. The lists found in the literature are not comprehensive. The problem is complicated by differences in terminology among the lists; in different lists the same name is used for different methods,

and different names are used for the same method. Methods listed separately by one writer are combined into one method by another. There is also confusion between methods for population prediction and methods for school enrollment prediction. An attempt is made here to describe some of the commonly named methods of enrollment prediction. The list is an incorporation and expansion of the classifications given by Peters (1969), Griffith (1964), MacConnell (1957), Strevell (1952), and the American Association of School Administrators (1947). One dimension for describing the methods is the amount of opportunity the forecaster has to depart from past trends in his forecasts. Methods which reply solely on data from past trends are referred to as projection methods; methods allowing deviation from past trends are called prediction methods. These terms were chosen by the investigator for convenience in the present study; some demographers use the terms somewhat differently (Metropolitan Area Planning Council, 1968:2 and Isard, 1960).

The methods which may be called projection methods rely on the trends of past enrollment figures. One method is that of fitting an equation to the curve of historical enrollment data ("projections by growth curve": Griffith, 1964:33-34). Collins and Langston (1961:10-12) listed three types of trend projections: (1) straight line, (2) average percent of increase,

and (3) average numerical increase. The straight line method consists of graphing the enrollment data for a grade or group of grades over a period of years and graphically projecting a straight line to determine projected enrollments. The average percent of increase, or geometric ratio method, consists of projection by applying the average annual percent increases in enrollment in a grade or group of grades. The average numerical increase, or arithmetic ratio, is a similar procedure. The average gain in numerical enrollment, rather than the average percent of increase, is applied to enrollment in successive years in the projection.

A method of forecasting school enrollment from total populations assumes that an observed ratio between total population and school enrollment that has existed in the past will also exist in the future; the ratio is then applied to an estimate of future population (American Association of School Administrators, 1947:55; Griffith, 1964:32-33). A similar method is described by Strevell (1952:37-38); it allows for a projection of the ratio rather than assuming a constant ratio. Straight lines fitted by the least squares technique to historical ratios of school enrollment to population have been used to project United States school enrollment (Simon and Fullam,

1968:89). Methods employing ratios of enrollment to population are presented here as projection techniques since the ratio is derived from past trends and the population itself is often projected from trends. Forecasts employing the ratio to a population which is predicted, rather than projected, are considered as a separate method; examples are discussed below under the names of "housing projection techniques" and the "land saturation method."

A more refined use of historical data is the "census class projection" described by Strevell (1952:35). It employs historical percentages of each census class enrolled in school, with a census class being defined as a given age group in a given year. Historical "migration ratios" are calculated by comparing each census class to the class one year younger the previous year and by comparing the age six census to births six years previous. Averages of these ratios over several years of experience are applied to census classes to simulate their advancement through the years of the projection. Historical percentages of each census class enrolled in school are then used to obtain enrollment projections. In this case, it is clear that both the enrollment percentages and the base population are projected rather than predicted.

Probably the most widely used method is the "percentage of survival" technique. The procedure involves percentage of survival ratios analogous to the migration ratios used in the census class projections. However, since the percentage of survival method employs grade enrollments, rather than census class counts, it is not necessary to apply percentages of enrollment in school. The basic method employs birth and enrollment records. The ratio of first grade enrollment to resident births six years previous is found by averaging the ratio over several years' experience. The ratios of the enrollment in grade g in calendar year z to the enrollment in the grade $g + 1$ in calendar year $z + 1$ are also averaged across several years of experience. Enrollments are projected by applying survival ratios to present enrollments to obtain projected figures for the first year and by applying the ratios to projected figures to obtain projection figures for successive years. The basic assumption is that the net effects of all factors which influence survival rates from birth through grade 12 will be the same for the projected period as they were during the period of experience used as a basis for projections (Peters,

1969:1).

The percentage of survival method has several other names: "forecasting by analysis" (MacConnell, 1957:30-31), "rentention ratio projection" (Strevell, 1952:32-36), "survival rate projections" (Griffith, 1964:38), and "percentage of rentention" (Greenawalt and Mitchell, 1966:4-5). Examples of the use of the percentage of survival method include projections made by the Massachusetts School Building Assistance Commission (Greenawalt and Mitchell, 1966). An enrollment study of Lexington, Massachusetts, used the percentage of survival method for short-range projections (Metcalf and Eddy, Inc., 1968); a study of Watertown, Massachusetts, employed percentage of survival ratios (Hunt, 1967). The method has been used for projecting elementary and secondary enrollments for the state of Nebraska (Nebraska Co-ordinating Council, 1967); the Florida Department of Education has developed a computerized planning system which utilizes the percentage of survival method in the enrollment projections (Daniel, 1969). Projections of school enrollment in the United States as a whole have been made by the percentage of survival method (Simon and Fullam, 1968:95).

Brown (1961) described a modification of the percentage of survival method which he called the "corrected prmotion"

method. Brown's method includes an estimate of the number of new residents in each grade each year, i.e., the number of children who might be expected from any new homes that are built in the school district (p.41). The number of new residents expected is added to the number of children in a grade before application of the survival ratio. Brown did not discuss the possibility that this procedure would quite likely overcorrect for migration. If the derived survival ratios are to any extent based on the arrivals of new residents, the additions of predicted new residents before application of the ratios would be an overcorrection.

In the study of Pittsburgh, enrollment projections were adjusted after percentage of survival rates were applied; additions and subtractions of students in specific grades and years were made if a housing development or housing removal project were planned for the district (Center for Field Studies, 1966). This procedure is subject to the same kind of criticism as Brown's corrected promotion method unless it can be assumed that the survival ratios do not project trends in housing development or removal. Slight adjustments were also made to the survival rates themselves to obtain high projections as well as most probable projections.

Survival rates were adjusted in some districts for the beginning of the projection period; other adjustments were made for the second five years of the ten-year projection period. However, the report does not contain an explanation for the rationale or amount of the adjustments.

Marshall (1968) used percentage of survival predictions adjusted for changes in trends. He made adjustments by selecting survival ratios, but he did not explain his method of selection. Similar kinds of adjustments were made in a study of Hartford; it is again unclear how the amounts of adjustment were determined (Center for Field Studies, n.d.).

A sophisticated technique for modifying U. S. Census state population projections to obtain enrollment projections for individual counties is the "multiple regression equation or cohort-ratio" approach (Jaffe, 1968). Independent variables in the model are school enrollment variables such as enrollments in groups of grades in the county; dependent variables are county statistics such as resident births, resident deaths, retail sales, number of households, income, and registered vehicles. All independent and dependent variables are expressed as the ratio of the county's share of the total state figure

in a given year to its share the following year. Multiple regression equations are developed, each relating one dependent variable to the independent variables, averaged across years for which the ratios are obtained. Dependent variables are projected by fitting statistical trend functions and extrapolating. The multiple regression equations using projected dependent variables are solved to obtain the projected cohort ratios of the independent variables, which in turn are used to obtain projected shares of state enrollment which the county enrollment will comprise.

A method in which the historical trends used in the projection are those of another community is "projection by analogy" (Griffith, 1964:34). An attempt is made to locate a community which has had a growth pattern similar to the community under study, but which is now larger in population and public school enrollment. Insofar as possible, social and economic conditions should be similar. Enrollment figures or rates in the comparison community are used to predict the future enrollments of the community under study.

Methods of forecasting which are at least in part predictive are "housing projection techniques" (Griffith, 1964:35-36; Strevell, 1952:37) and the "land saturation

method" (Peters, 1969:4-5). The predictive aspect of the housing projection techniques is the prediction of the number of households or dwelling unit types. Enrollments are calculated by applying the average enrollment per household or dwelling unit type. The average enrollment rate is most often based on historical data. After obtaining an estimate of future numbers of dwellings and number of children per dwelling, a committee making enrollment forecasts for Lexington, Massachusetts, applied a percentage enrollment figure calculated from historical data (Population and Housing Committee, 1966). The "land saturation method" described by Peters (1969:4-5) is used to predict the growth of the population on the basis of anticipated use of available land for additional industrial and residential buildings; it is not made clear exactly how the population figures are translated into enrollment figures.

A method which is designed specifically for enrollment forecasts rather than population forecasts is referred to as the "multivariable method" (Johnson:1965) or presented without a specific name (Center for Field Studies:1953). It can be used with input which is drawn from past trends

or that which is anticipatory of future trends, with the latter type of input being the most common for the multivariable method. The basic model employs estimates of each of the major factors affecting school enrollments and applies the factors separately to previous enrollments to derive projected enrollments. The exact statement of the model varies; a typical model is given by Johnson (1965:186-187):

A. (Survival births) + (Preschool net migration) + (Retentions in grade 1 previous year) - (Non-public school enrollment grade 1 this year) = (Estimated public school enrollment in grade 1).

B. (Public school enrollment grade 1 previous year) + (Non-public school enrollment grade 1 previous year) + (Net migration) - (Non-public school enrollment grade 2 this year) - (Retentions grade 1 previous year) + (Retentions grade 2 previous year) - (Dropouts grade 2 previous year) = (Estimated public school enrollment in grade 2).

A report by Arthur D. Little, Inc. (1966) on Quincy, Massachusetts, schools included a similar model based

on historical data. Johnson (1965) modified the calculated historical trends to obtain input for his model. After calculating the projected net migration figure for women of child-bearing age, he reduced the calculated figure somewhat to account for his assumption that the rate of population growth was declining slowly in Englewood (p.171). He also made the assumption that 1960 fertility rates would decline slightly during the next few years (p.172), and he made "intuitive judgments" about the changing character of pre-school migration (p.177). In estimating school-age migration, he hypothesized that most of the out-migration during 1960-64 had occurred in 1962 and 1963; his "most probable" estimate of future school-age net migration reflected a smaller loss of school-age population than an average of the experiences of the last four years would have projected. He made alternative predictions based on different assumptions about future private and parochial school enrollment.

A variation of the multivariable model was used in a study of the greater Corning area of New York State (Center for Field Studies, 1954). Survival ratios were used, but adjustments were made for anticipated changes in the housing situation (p. Ap.16). Included in adjustment of past trends

for nonpublic school enrollment were the two new rooms planned for the parochial schools. Retention and dropout rates were projected.

Another example of the multivariable method in which some of the variables were predicted and some projected is a study of Arlington, Massachusetts (Center for Field Studies, 1953). Changes in trends in construction rates and available land were considered in predicting migration ratios. Projections were made for variables such as dropouts.

A multivariable study of Salem, Massachusetts (Center for Field Studies, 1956) is exemplary in its reporting of the considerations involved in adjusting for anticipated changes in trends:

> The computation...was adjusted to account for the opening of the housing project at Rainbow Terrace since the in-migration of this period cannot be expected to be repeated each succeeding five-year period [p.23].
>
> A study of the economy of Salem, its housing, and its probable future in the area would seem to indicate that the average migration of the last fifteen years would be more typical of the future than the average migration of the last five years. During these years there has been an unusual out-migration apparently attendant upon the closing of the Pequot mill [p.24].
>
> The Diocesan Superintendent reports no knowledge of any plans to expand parochial school facilities in Salem at the present time [p.26].
>
> The Diocesan high school, to be located at Peabody, is planned. If this high school is built, Salem high school pupils might attend.

> It is estimated by the Diocesan office that possibly 100 boys and girls from grade 8 of Catholic schools would attend the freshman year at the new high school [p.26].

A study of Brockton (Center for Field Studies, 1964) is another example of the multivariable method using input that goes beyond extrapolation:

> The somewhat arbitrary decision to use a base index of 850 for projecting future births is justified by the following arguments:
> 1. It is consistent with the prediction of slowly rising fertility ratios during the 1960's.
> 2. It seems to parallel the allocated birth pattern observed in Brockton since 1960 [p.A-6].
>
> Rather than show the details of this net migration estimate, we will explain the key assumption which was used. Namely, approximately two-thirds of the female net migration in Brockton between 1950 and 1960 is likely to occur during each five-year period between 1960 and 1970. The reason for using this fraction of two-thirds is two-fold. First, two-thirds of the building permits registered in Brockton between 1950 and 1960 were issued during the last half of the ten-year period. Second, there seems to be an indication that the issuance of building permits (and hence, the net migration figure) will not increase much over the 1955-60 level [p.A-7].
>
> After talking with the principal of each local private and parochial school, a projected yearly capacity was determined for each building. Since the demand for Catholic education appears to exceed the current capacity, the projected enrollment is simply the estimated capacity of these individual schools [p.A-10].

Griffin and Schmitt (1966) proposed a model similar to the other multivariable models.

There are many ways in which the estimates for the variables in the model may be obtained; some of the methods may be unique to an individual school system. To some extent, other methods of population and enrollment forecasting can be incorporated into the multivariable method. For instance, projections of housing trends could be used in the estimate of pupil migration. Estimates may be made by those persons most familiar with each of the separate variables or by one person who gathers information from many people and sources. General guidelines for forecasting may be found in the literature on school enrollment prediction. Scammon (1962:39-41), Brown (1961:13,17, & 41), and Greenawalt and Mitchell (1966:24-31) suggested as sources the following records, agencies, and persons: U.S. Census publications, local school censuses, birth records, marriage licenses, building permits, utility company records and projections, mail route changes, U. S. Office of Education, National Education Association, Chambers of Commerce, planning boards, zoning commissions, school enrollment reports, building and occupancy reports, tax records, real

estate developers, and government and news publications which discuss general business conditions and the activity of home and industrial construction industries.

The multivariable method was chosen as the basis of the procedure for producing probabilistic enrollment output. The literature on population and school enrollment forecasts gives suggestions of criteria for selection of a model. One criterion is that the model not restrict the forecaster to mere projection of past trends if he feels that there is an indication that the trends may take a new direction in the future. Evidence is seldom given in enrollment studies for the assumption that enrollment trends are stable and are likely to remain so; the assumption of continuation of past trends seem most often to be an option for simplicity. It is this type of assumption which is criticized by many demographers (Isard, 1960; Rosenberg, 1968:3) who emphasize that such an assumption is often unjustified. Isard (1960) said that there is no assurance that a graph or function that fits past data will adequately describe the future pattern. Gottlieb (1954:68, 110) stated that enrollment forecasts must not simply project past trends, but must take into

account the probable effects of political, social, and economic factors.

The word "projections" in this study is reserved for those studies which project trends for such reasons as expediency and simplicity. An assumption of continuation of past trends is just as adequate as an assumption of change if it is based on the same careful reasoning about future developments. This type of reasoning is present in a study of Boston:

> Since the basic curve assumes that the influences affecting migration remain the same, evidence of the reasonableness of this position was sought. Project directors of the Boston Redevelopment Authority, settlement house directors, and welfare agency officials were among those asked whether to their knowledge any factor, aside from urban renewal, existed which would alter the population trend for the health and welfare area under discussions. Invariably, the response was the same: i.e. that except for the impact of the urban renewal program, public housing, and the growth of Negro population, the present population trends would continue [Sargent, 1962:A-8].

Probable effects of these three factors were examined.

Any method can be adjusted to account for trend changes; the multivariable method, however, seems to be the most adaptable to these adjustments. The other methods discussed above involve, to a greater extent than does the

multivariable method, summary coefficients or statistical relationships among variables. For example, the percentage of survival method is based on the assumption that enrollment changes can be expressed as the relationship of enrollment in one year to enrollment in another year; the trend projection techniques and the ratio of enrollment to population techniques also assume stability of relationships among variables. Two ways of adjusting these types of methods for trend changes are to adjust the coefficients or statements of relationship or to adjust the enrollment projections themselves. Brown (1961) attempted the latter; the problem with his method, as discussed above, is that he risks incorrect adjustment for the migration rate. It is difficult to adjust the enrollments after a coefficient, equation, or graphic projection has been applied. In adjusting for, say, changes in migration trends, the summary coefficient gives no indication of the migration trend which it tacitly projects and it is, therefore, difficult to determine the amount of adjustment necessary. To adjust the coefficient itself is equally difficult; the contribution that migration rates make to the value of the coefficient should be determined.

Methods which allow separate estimation of at least one variable, such as the housing projection techniques, are somewhat more flexible; the enrollment ratios used in these methods, however, combine estimates for variables such as birth rate, dropouts, retentions, and nonpublic school attendance.

Analyzing enrollments by separate variables, as in the multivariable methods, simplifies adjustment of single variables (Center for Field Studies, 1956; Whitla, 1954). Of course, the multivariable method does not consider separately every possible variable affecting school enrollment; for example, it treats migrations as a single variable, whether the migrations are the result of a new industry, a housing boom, or a highway. For generalizability and manageability, the multivariable method does group variables to some extent; however, the forecaster is encouraged to consider separately the various factors in the school system which relate to migration and the other variables.

Another argument can be made in favor of the multivariable approach: it is more adaptable to probabilistic modifications of the input. Just as any other method could be adjusted in some way for predictive input, any method could be adjusted for probabilistic input. It seems, however,

that the forecaster has a more rational basis for placing probability limits around estimates for a single variable such as migration than around coefficients or other statistical summaries. Of course, if he were willing to make the necessary assumptions, he could place probability limits around a coefficient by analyzing the variance of past data. An example is given by Peters (1969). He computed not only the means of historical survival ratios, but variances as well. Confidence intervals of projected enrollments are derived from the use of the variances of the survival ratios.

A third reason for the desirability of easily adjustable predictions is the opportunity for precise adjustment of the long-range predictions after the short-range predictions for the separately estimated variables have been validated by experience (Center for Field Studies, 1956:22).

Another advantage is that an easily adjustable method can be used to simulate the results under different assumptions about the various variables. For example, the effects of a contemplated change in the policy for retaining students can be studied. The multivariable model, even without the addition of Monte Carlo simulation, can be considered to be

a type of simulation. Thus, it has some of the advantages of the simulation model, one of which is the possibility of simulating the results of a policy decision rather than relying on costly trial and error (Malcolm, 1958:57). An example of the use of a multivariable model in this way is a project undertaken by UNESCO (1966).

The model chosen should not only be easily adjustable for future trends and probabilistic input and output; it should be demonstrated to be demographically sound. However, school enrollment forecasting techniques have not been rigorously examined. There is a record of only a few studies directed toward an evaluation of the various types of forecasting methods and these evaluations have been far from satisfactory (Jacob S. Siegel, 1953). What evidence there is seems to indicate that all methods are susceptible to startling errors under certain conditions. In addition, several authorities have despaired over the seeming impossibility of producing accurate population forecasts for small areas. Even national and state population forecasts, considered surer than those of small areas, have in the past demonstrated gross errors. In 1930, W. S. Thompson and P. K. Whelpton predicted that the 1960 population of the United States would be between 137.9 million and 167.3 million; the 1960 census figure was 179.3 million (Greenawalt

and Mitchell, 1966:5).

Greenawalt and Mitchell (1966) examined the accuracy of the percentage of survival method using predictions for the years 1952 through 1959 made by the Massachusetts School Building Assistance Commission for 242 towns and cities in Massachusetts. It was assumed that a forecast which predicted enrollment within plus or minus 10 percent of the actual enrollment was "accurate." Of the 242 predictions studied, 149 were found to be inaccurate by this definition (p. 8). They concluded that the percentage of survival method is most likely to be in error in fast growing communities (p.15).

Experience has shown graphic techniques not to be very accurate methods of forecasting (American Association of School Administrators, 1947:55). Arithmetic and geometric progressions have been shown to be "surprisingly accurate," but only because the projections were continually reassessed to account for current trends (Center for Field Studies, n.d.:41).

Larson and Strevell (1952) reviewed 31 enrollment forecasts made in twelve states during 1930-1952 by local boards of education, private survey firms, taxpayers' associations, the U. S. Office of Education, and schools

of education and extension services of universities. The first two of the three methods reviewed are similar to the ratio of enrollment to population methods. In the first method, the population is estimated by statistical projection of past population. In the second, the population is estimated by consideration of community factors such as numbers of gas and electric meters, water meters, and telephones. It is not clear in the description whether these data are to be used for projection or for prediction; in the latter case, the method would be similar to the housing projection techniques discussed above. The third method discussed is similar to the percentage of survival method. None of the three techniques were shown to be clearly superior in terms of gross error. Although no statistical tests were performed, the report contained the median percentages of error for forecasts of one to five years; the percentages were, respectively, 6.2, 9.9, and 7.7. Forecasts made by the first method were more likely to be overestimates.

According to Whitla (1954), the multivariable method is "most accurate" in estimating enrollments. He reported a study in which percentage of survival forecasts were not as accurate as multivariable forecasts. The study was pre-

pared by the United States Bureau of the Census for forecasting statewide school enrollments.

The review of the literature on enrollment forecasting techniques seems to justify the choice of the multivariable method for prediction and as the basis of the probabilistic technique developed in this study. The form of the multivariable model used in the present study for grades 8-12 is diagrammed in Figure 2.1 (p.38). The form of the model for the other grades is similar. The forecaster places probability limits around the separate variables in the model. Specifically, the forecaster is asked to give each variable a high estimate, a most likely estimate, and a low estimate, with the high and low estimates representing the limits of the 98 percent confidence interval. The problem is accounting for this probabilistic input in the prediction output.

Similar problems have been conceived in terms of Markov chains. The model would have to be modified somewhat to be considered in these terms; for example, migration figures would have to be considered in terms of probabilities that a student in a certain category would migrate. A finite

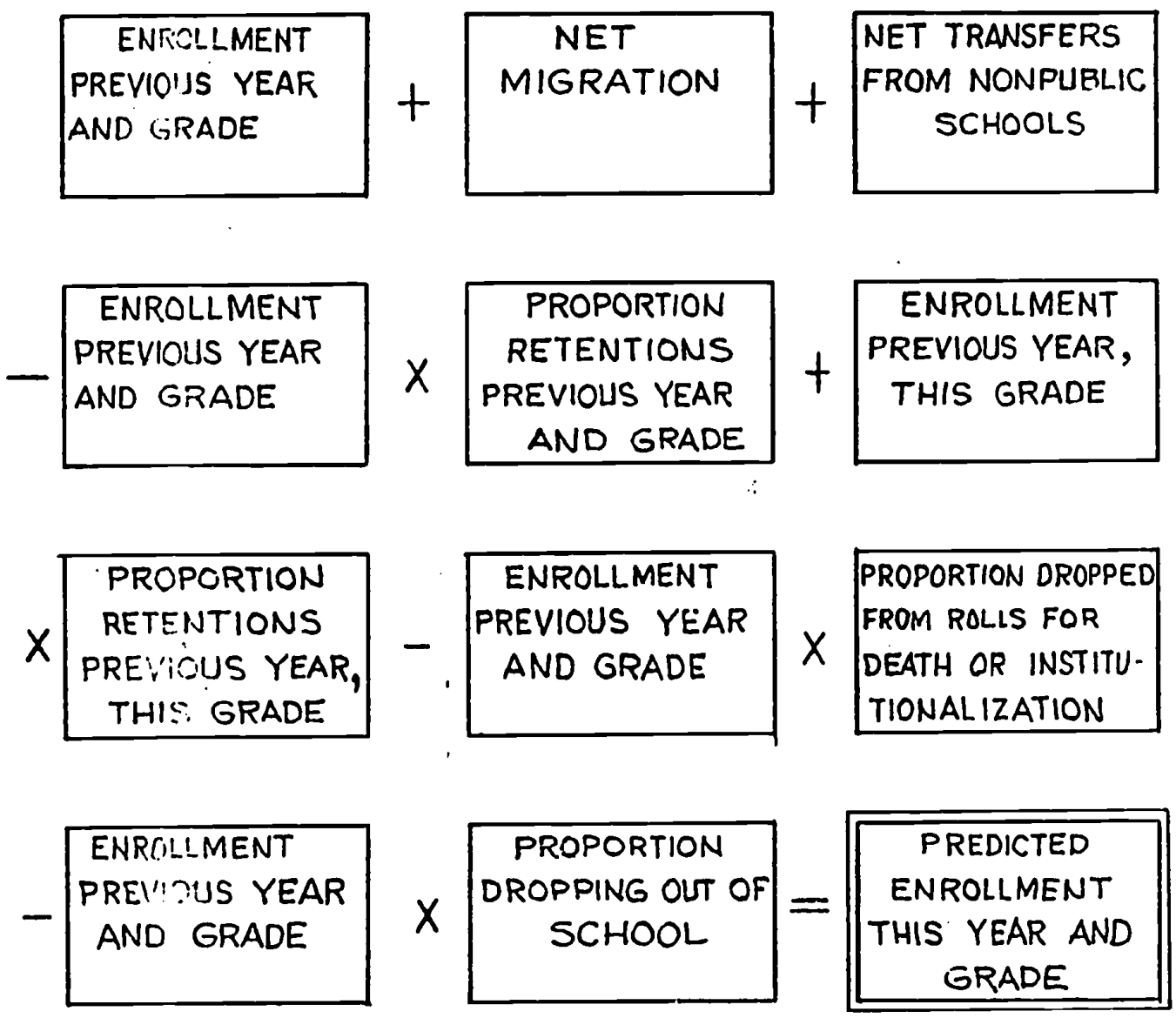

Fig. 2.1. Multivariable model used in the present study.

Markov process, or Markov chain, is a multistage stochastic process such that the probability of a process being in any one of a finite number of states at time $t + 1$ is conditional on the state the process is in at time t and the matrix of probabilities of moving among the states (Burford, 1966:5; Bartos, 1967:31). In terms of the multivariable model, the states might be categories such as first grade student, second grade student, dropout, or deceased. There are problems, however, in conceptualizing the problem in terms of Markov chains. One of the basic assumptions of Markov chains is that the transition probabilities would be the same for each year of the prediction. A second restriction is that it must be possible to obtain estimates of probabilities of moving from any given state in one period to any other state in the next period. It is difficult to express births and in-migrations as a percentage of a set of possible in-migrations and births. The logical method of representing in-migrations would be to have a state called "world" (Bartos, 1967:135), but to estimate in-migrations in a school system as a percentage of the world population or even of the United States or a single state's population is infeasible.

An example of a computerized application of a modified

Markov chain to school populations is Dynamod II (Zabrowski, 1969). The assumption of stable transition probabilities in this situation is not serious because the purpose of the model is not to predict new trends in school enrollment, but to provide educational planners with information on the impact on educational populations of proposed policy changes or of sudden shifts in the structure of the ec cational system, by varying such factors as dropout rate or teacher-pupil ratio. Migrations were not included in the model since the study involved an analysis of national enrollment figures, but including births necessitated modification of the Markov chain model. The numbers of people within the system moving among states were calculated by transition probabilities, and estimated births were simply added to categories when applicable. Thus, Zabrowski circumvented the problem of transition probabilities involving the state representing birth.

A similar study was done in Norway (Thonstad:1967) and was based on transitional probabilities for such states as grades, schools, deaths, dropouts, and graduates for the purpose of determining the educational distribution of the population to which the present propensities were leading.

But again in this use of Markov chains, predictions of changing enrollment trends were not the purpose of the study. Thus, Markov chains have not been applied in these instances to essentially predictive problems and the underlying assumptions of Markov chains limit their usefulness for solving the problem of calculating probabilities for predicted enrollments.

Before considering Monte Carlo simulation as a solution, it is necessary to consider the possibility of straightforward mathematical solution of the model as diagrammed above (Figure 2.1). "Every Monte Carlo computation that leads to quantitative results may be regarded as estimating the value of a multiple integral [Hammersley and Handscomb, 1964:50]." The multivariable model as modified in the present study to accommodate probabilities is built with the assumption that the probability distributions around the separate variables can be considered to be normal. The model is similar to the linear combination of n independent random variables which are normally distributed. Thus, for normally distributed variables in the model

$$Y = C_1X_1 + C_2X_2 + \ldots C_nX_n ,$$

the distribution of Y is normal and has a variance given by

$$\sigma_y^2 = c_1^2\sigma_1^2 + c_2^2\sigma_2^2 + \dots c_n^2\sigma_n^2$$

(Hays, 1963:234, 236). For a normal distribution of enrollments with known variance, it would be a simple task to express the various probabilities. However, the present model does not fit the pattern exactly; the model is not simply a linear combination of the variables. As shown in Figure 2.1, some of the variables are multiplied; for example, the model includes the product of rentention rate and enrollment the previous year. Thus, the variances of the variables cannot be added to determine the variance of enrollments, and the final distribution of enrollments is not necessarily normal. The model could be adjusted by calculating the variance of the products and then calculating the linear combination of single variables and products of variables. To calculate the variance of a product of variables X and Y, the following formula is used:

$$\sigma^2(XY) = \sigma^2 X \sigma^2 Y + \bar{Y}^2 \sigma^2(Y) + \bar{X}^2 \sigma^2(X) - \bar{X}^2 \bar{Y}^2 ,$$

where $\bar{X}$ is the mean of X and $\bar{Y}$ is the mean of Y (derived from Kendall & Stuart, 1963:232-233).

However, the use of the above adjustment assumes that all of the variables are independent. This assumption is violated in that previous enrollment is one of the single

variables and it also appears in one or more of the products of variables. For example, the product of a retention rate of .10 and a previous enrollment of 1964 is subtracted from the previous enrollment of 1964; in this case, the two enrollment variables are obviously non-independent.

A further complication is that it is more difficult to assume a normal distribution of calculated enrollments. Without an assumption of normality, it is more difficult to derive probabilities from knowledge of the variance.

Monte Carlo simulation is proposed here as a satisfactory way to combine the probabilistic input to obtain the probabilistic output. A computer simulation model is a logical-mathematical representation of a concept, system, or operation programmed for solution on an electronic computer (Martin, 1968:5). Solving a problem by the Monte Carlo method amounts to submitting the problem to a roulette wheel (McCracken, 1955). Extending this analogy, the compartments of the wheel are labeled in such a way as to reflect the probability distribution. For instance, if a student had a 95 percent chance of passing his course-work, one out of 20 compartments would be designated "fail"; the others would be designated "pass." A spin of the roulette wheel would

produce the "iteration-specific" value of the variable. Spinning the wheel again and again to produce more iterations would result in a distribution of iteration-specific values. In this case, it would be a distribution of values with approximately 95 percent of the values representing "pass." The distribution of the outcomes in this problem is easily predicted. The solution to the enrollment problem proposed here is analogous to spinning the wheel several times for each iteration, with one spin for each variable contributing to the enrollment prediction. In this type of problem, the outcome is not so obvious.

In a Monte Carlo solution, one or more of the variables depend upon chance parameters whose values are randomly selected from probability distributions. Because of this random feature, the outcomes in a Monte Carlo solution usually differ for repeated runs with the same input values; to produce statistically significant results, replications are required with the same inputs (Martin, 1968:33). In summary, Monte Carlo simulation involves random sampling from probability distributions to determine iteration-specific outcomes; the results of a number of iterations form the solution of the problem.

Although the mathematical techniques used in the Monte Carlo simulation had been used before, particularly in deterministic problems, the name and popularization of the technique dates from 1944 (Hammersley and Handscomb, 1964: 6). Research on the atomic bomb during the Second World War involved a study of random neutron diffusion in fissile material. The scientists had basic data such as the average distance a neutron of a given speed would travel before collision with an atomic nucleus, the probability that the collision would result in a neutron bouncing off rather than being absorbed, and the amount of energy the neutron was likely to lose in the collision. However, it was impossible to sum all of these probabilities in a mathematical formula. John von Neumann and Stanislas Ulam suggested a solution which was given the code name "Monte Carlo" (McCracken,1955:90).

The Monte Carlo approach to the solution of such a problem consists of pretending to trace the life histories of a large number of neutrons. At each decision point in the history of each neutron, a random number is selected to determine which outcome occurs, in keeping with the known probabilities of occurrence. Through the accumulation

of a large number of such histories, it is possible to estimate the percentage of neutrons which will terminate in each of the final possible outcomes (Hammersley and Handscomb, 1964:6).

Monte Carlo simulation can be used to solve both deterministic and probabilistic problems (Meyer, 1954:2). Problems are classified as deterministic or probabilistic depending on whether they are concerned directly with random processes or involve the assumption of a random process for the purpose of obtaining an approximate solution to the problem. The study of the behavior of neutrons was essentially probabilistic. An example of a deterministic problem is the approximation of area under a curve, not by calculus, but by recording the percentage of occasions on which a pin thrown randomly onto a graph falls inside, rather than outside, the curve.

The present Monte Carlo simulation is analogous to the neutron behavior study in that the enrollments are considered to be stochastic, as were the neutron behaviors. A random number is used to determine iteration-specific enrollment variables, the combination of which represents an iteration-specific predicted enrollment. The model differs from stochastic Monte Carlo models such as the neutron simulation

in that the factors assumed to be probabilistic are the values of the variables such as migration rather than the behaviors of individual students or neutrons. The proposed use of Monte Carlo for predicting school enrollments is more closely analogous to problems in operations analysis in which the random variables may be, for example, the estimated time needed to accomplish certain tasks. McCracken(1955) discussed the use of Monte Carlo in a wood-working shop problem to determine how the work of the shop should be scheduled to yield the greatest production, considering a number of variable conditions. If one knows that a certain job would take from 12 to 16 minutes, respectively, he could simulate the operation by random sampling. In the school enrollment problem, the values of variables such as migration are analogous to the time variables in the shop problem.

According to Malcolm (1958:57), simulation has the advantage of being easily understood because it is relatively free of complicated mathematics. Also, a mathematical solution may be unavailable or too complex to apply. Orcutt (1962:101) stated that Monte Carlo simulation enables the user to introduce into his model interactions, variables, non-linearities, and stochastic considerations that he might

not be able to introduce using a mathematical method.

Monte Carlo simulation has been applied to problems in demography. Sheps applied Monte Carlo techniques to construct stochastic micro-models of demographic behavior (Demeny, 1964:70). Part of Sheps' model was described by Beshers (1964):

> She puts each individual women of the cohort stochasically through a number of contingencies, year by year. First, she ascertains how long each women is likely to live; when she is likely to marry for the first time; and when she is going to become sterile. Once a women is married and fertile, the program ascertains, by the same kind of stochastic process, whether or not she is a family planner. Then her history through child-bearing is determined from her age, parity, and the other contingencies that occur [p. 72] .

A team led by Orcutt (1961:285-350) conducted a socio-economic simulation of the United States using Monte Carlo simulation. The simulation resulted in a distribution of people among various demographic categories.

Other kinds of simulation have been used to study school enrollments. Members of the UNESCO staff developed a system simulation for Asian countries (1966). Mathematical models for enrollment projection which do not include Monte Carlo simulation have been developed for Sweden (Forecasting Institute, 1967) and Great Britain (Armitage and Smith, 1967).

The potential value of Monte Carlo in school enrollment studies has been recognized, but not yet applied. Armitage and Smith (1967) outlined the reasoning:

> Even at this stage, our model calculations will still be purely deterministic, i.e. everything happens exactly as specified by the equations. . . .If we are to make some assessment of the value of deterministic calculations, then it is convenient to regard the transition proportions as probabilities. . . .The model will then consist of simultaneous multinomial distributions and while the explicit treatment of such a system would be intractable, it will be possible to carry out Monte Carlo simulation calculations in which the values of all (variables) ... are found by sampling from multinomial distributions. The introduction of random variation in this way is one line of attack upon the 'noise' problem, i.e. the distortion of the signal of the deterministic calculation by the presence of variation, the fact that things do not happen exactly as expected. The deterministic calculation can still be regarded in the traditional way as providing 'the best estimate' of what is expected to happen; but these simulation calculations will give some indication of how far we are justified in placing our faith in these 'best estimates' and acting upon them [pp.184-185] .

Griffin and Schmitt (1966) detailed a model for using Monte Carlo simulation in enrollment predictions. However, the actual simulation was not performed.

The present study includes the performance of the simulation, as well as basic changes in the Griffin-Schmitt model and tests of validity and reliability not included in their design.

Chapter III

Design of the Study

The major purpose of the study was to develop a satisfactory method of enrollment forecasting to produce multiple figure predictions with associated probabilities. A review of the literature indicated that a Monte Carlo simulation using the multivariable method as a basis for the simulation might well be a satisfactory solution. Thus the first major objective of the study was to develop a system of computer programs for performing Monte Carlo simulations employing the multivariable model. The first objective involved the following steps:

(a) collection of data on which to test the simulation,

(b) adaptation of the multivariable model to the present purposes,

(c) writing the computer programs for the simulation, incorporating previously written programs to generate random numbers,

(d) performing the simulation using the sample data, and

(e) preparing instructions for future users of the simulation method.

The second major objective was to investigate the predictive validity, reliability, and concurrent validity of the method using statistical tests with sample data.

Data needs for the present study were somewhat unique in that the required data were not measures on variables for various populations, but predictions of school enrollments. Historical predictions, that is, predictions which were made in some previous year, were needed for the test of predictive validity. The reason, of course, is that historical predictions can be tested for their accuracy of prediction for the intervening years.

The data chosen to represent the multivariable model were the basis of an enrollment prediction study for Brockton, Massachusetts, made by the Center For Field Studies (1964: Appendix). The data had to be modified somewhat; the most significant modification of the data made by the present investigator was the addition of high and low estimates for the variables, which were given only one estimated value by the Center for Field Studies. The data from this particular enrollment study were chosen after an extensive search of the literature and over thirty interviews with educational consultants, university personnel, state department of education

personnel, school administrators, city planners, and school committee members. The search revealed that the multivariable method was not among the most common methods presently used. However, the method was used in several studies by the Center for Field Studies of Harvard University Graduate School of Education. One of the studies, that of Brockton, Massachusetts, seemed to be most adaptable for the present purposes; an example of the multivariable method with "high" and "low" estimates for the variables was not available.

Brockton is a city in southeastern Massachusetts, twenty miles from Boston, with a 1965 population of 83,499. During the decade of 1955 to 1965, the population of Brockton increased by 20,871, with an estimated excess of births over deaths of 8,167 and an estimated net in-migration of 12,704 persons. Brockton is an industrial city with shoe manufacturing the predominant industry. The median income of Brockton families in 1960 was $5,914, somewhat below that for the state as a whole (Massachusetts Department of Commerce and Development, 1967).

The Brockton data provided predictions for each of the variables and for total enrollments for grades 1 through 12 for the academic years 1964-1965 through 1975-1976. Since the

Brockton study utilized fall enrollment figures and the computer programs developed in this study were designed to predict fall enrollment, a prediction for the academic year 1964-1965 is considered a prediction for the year 1964.

Besides the addition of high and low estimates for variables, the data provided in the Brockton enrollment study had to be modified to fit the specifications of the particular multivariable model developed for the present study. In Figure 3.1 (page 54) is the model used by the Center for Field Studies to predict grade 8 enrollment; in Figure 3.2 (page 55) is the model used in the present study. Accompanying both models are the data for predicting grade 8 enrollment in 1964. Variations of the latter model for predicting other grades include use of birth data in predicting grade 1 and omission of the dropout variable below grade 8; the model for grades 9-12 is identical to the model for grade 8. The model developed by the Center for Field Studies is referred to as the "Center" model and the one modified by the present investigator is the "modified" model.

PUBLIC SCHOOL PREVIOUS YEAR GRADE 7 (1205)	+	PRIVATE & PAROCHIAL PREVIOUS YEAR GRADE 7 (156)	=	TOTAL PREVIOUS GRADE 7 (1361)
TOTAL PREVIOUS GRADE 7 (1361)	+	NET MIGRATION (3)	=	POTENTIAL GRADE 8 THIS YEAR (1364)
POTENTIAL GRADE 8 THIS YEAR (1364)	−	PRIVATE & PAROCHIAL THIS YEAR GRADE 8 (157)	=	IN PUBLIC THIS YEAR 1207
IN PUBLIC THIS YEAR (1207)	−	RETENTION FROM PREVIOUS YEAR, GRADE 7 (36)	=	PROMOTED TO GRADE 8 (1171)
PROMOTED TO GRADE 8 (1171)	+	RETENTION FROM GRADE 8 PREVIOUS YEAR (34)	=	PUBLIC ENROLLMENT THIS YEAR GRADE 8 (1205)

Fig. 3.1. Multivariable model for grade 8 used by the Center for Field Studies.(In parentheses are the figures used to predict 1964 enrollment.)

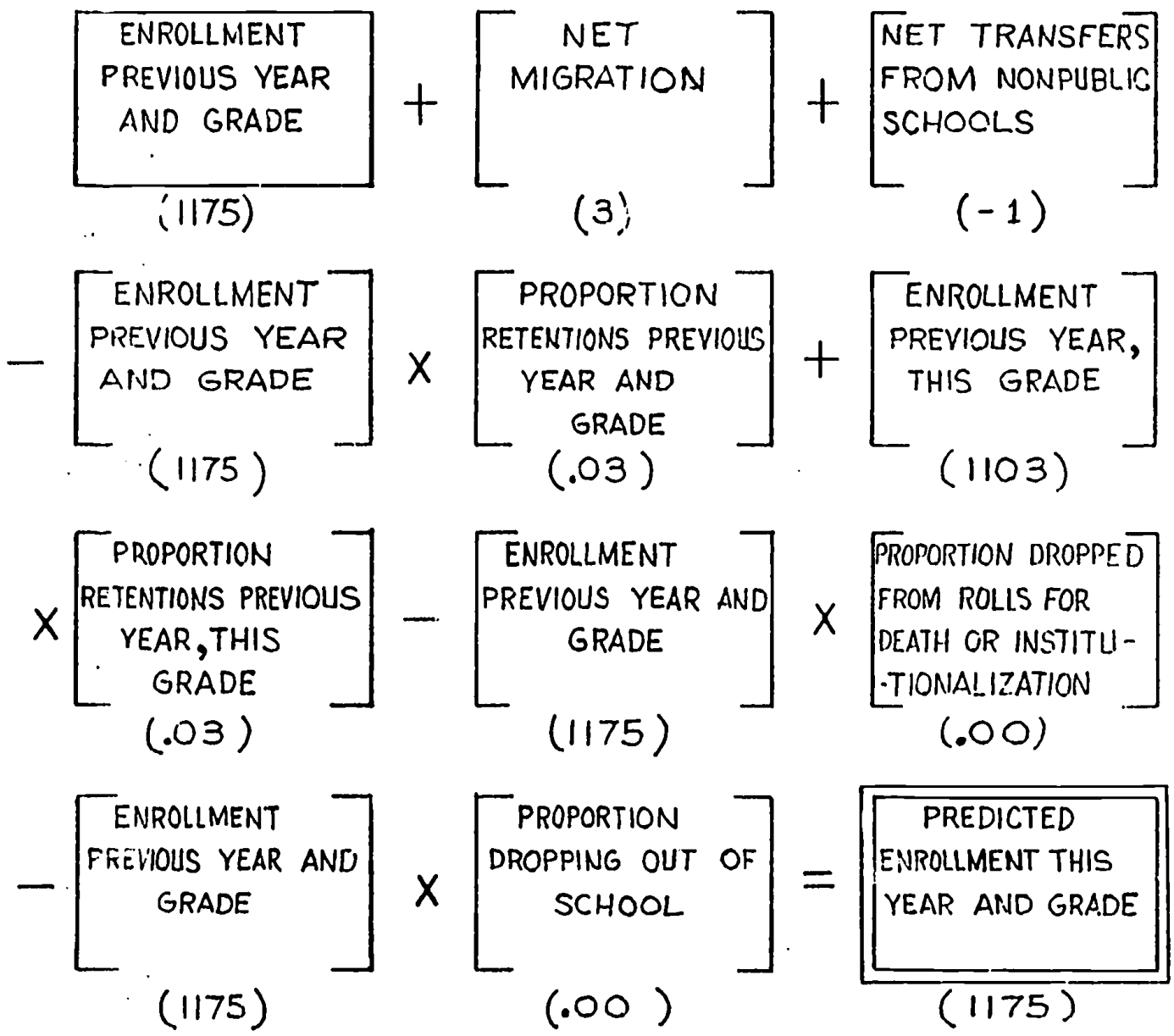

Fig. 3.2. Multivariable model for grade 8 used in the present study. (In parentheses are the figures for predicting 1964 enrollment without simulation.)

The modified model includes the dropout variable in predictions for grades 8-12; the Center model included it for grades 9-12 only. The dropout variable for grade 8 was used in several of the enrollment studies reviewed; in order to make the model as generalizable as possible, the variable was included in grade 8 of the modified model.

The modified model also includes rate of death and institutionalization as a variable at each grade level as recommended by Collins and Langston (1961:10). The Center model included deaths for only the first year of age. The modified model uses a variable for preschool deaths, which includes deaths during the first year. If a user wishes to estimate first year deaths only, as in the Center model, the preschool death variable may be used to record an estimate for first year deaths, and all values of the variable for school-age death or institutionalization may be set at zero.

It has been recommended (Center for Field Studies, 1954: App. 13) that age-specific birth rates be used when estimating future birth rates. Age-specific birth rates provide for the fact that women of certain ages have more children that do women of other ages; estimates of birth rate by this method take into account shifts in the age distribution of the female

population and birth rate for each of six age groups. The age groups used for age-specific birth rates are 15-19, 20-24, 25-29, 30-34, 35-39, and 40-44. The reports show a few births occurring before or after these ages. However, the numbers involved are too small to give significant results, and it is probable that some of these are due to mistakes regarding age (Whelpton, 1954:28). Births in Brockton were estimated by age groups by the Center's staff, but the Center model itself requires only that total births be estimated.

Similarly, one assumption on which the modified model is based is that the variables concerning first year deaths, retentions, student deaths and institutionalization, and dropouts, can best be predicted as proportions, rather than numbers, since the size of these estimates seems closely related to the size of the base population. The modified model requires estimates of these proportions; the Center model requires only estimates of the numbers of students involved, although these numbers were obtained in the Brockton study by projections of past trends in the proportions of students involved.

Both models have an advantage over some prediction models in that they require that migration and nonpublic school enrollment be estimated as numbers, rather than as percentages. An advantage of the multivariable method in general is that it does not confound variables to the extent that they cannot be accurately predicted; it would be unnecessary compounding of variables to estimate migration or nonpublic school enrollment as a percentage of, say, total school population (Center for Field Studies, 1954: App. 9). The assumption is that demographic data such as housing data are more accurately used to estimate the numbers of migrants than to estimate the percentage of migrants in total school population, which has other sources of variation in addition to migration.

Two variables in the Center model, private or parochial school enrollment for the previous year and private or parochial enrollment for the predicted year, are combined in the modified model to form one variable: net transfers to/from nonpublic schools. Besides simplifying the model, the combination makes it possible to collect much of the data necessary for the prediction from public school records. The nonpublic school variable can be adjusted in cases where there are children residing in the district attending school outside

the district and out-of-district children attending public schools in the district under study.

Another slight difference between the Center model and the modified model is that all net migrations after grade 1 in the modified model refer to in- and out-migration of public school students. The Center model allows for the estimation of total migration in the district for each grade, including migration of nonpublic school students. The reasoning behind the modification was to simplify the task and to make as much of the data as possible available from public school records. Some schools, of course, would have to begin keeping the necessary records if this prediction method were to be used.

Some interpolation of the Center data was necessary to obtain predictions of the number of women in each age group and the age-specific birth rate for each year of the simulation. The Center data provided estimated female population for the years 1965 and 1970 only; births were calculated for these years by applying age-specific birth rates, and the predicted births for the remaining years were obtained by interpolation. Since the modified model requires age-specific birth rates and migrations of women for each year, these variables had to

be interpolated from the 1965 and 1970 values. These interpolations were adjusted, as necessary, to coincide with the births as interpolated directly.

The Center data did not include a breakdown by sex for the predictions, but the proportion male for each estimated variable was needed to test the functioning of the part of the computer program designed to produce sex breakdown. Since no tests of validity for the sex breakdown were performed, a somewhat arbitrary estimate of the proportion male was satisfactory. The proportion male chosen for each of the variables was .516, the proportion male in the 1963 total public school enrollment (Brockton, Massachusetts, School Department, 1963).

A major revision of the Center data necessitated a recalculation of the predictions for the modified model. The year 1963 was used as the base year; that is, the historical enrollments by grade in 1963 were adjusted to become the predicted enrollments for 1964, the first year to be simulated. The historical enrollments recorded for 1963 in the Center's study were not the same as the 1963 enrollments recorded in the Annual Report of the School Department of Brockton (1964). Since historical data from the Annual Reports were used to

calculate the percentage of survival projections for use in the comparison of the predictive validity of the multivariable method with that of the percentage of survival method, consistency in the 1963 figures was necessary. Howard Johnson (1969), formerly with the Center for Field Studies, explained that the enrollments had been adjusted for factors such as special education students and homebound students. Since he was not able to state the exact means of adjustment, the decision was made to use the figures from the Annual Report for all 1963 enrollments.

Since the required high and low estimates were not made in the Center's study, these were set somewhat arbitrarily by the present investigator. However, guidance for these estimates was available; recent historical data for most of the variables were given in the Center's report.

The criteria for choosing high and low estimates for each of the variables are based on the assumption that the predicted values of the variables represent a probability distribution in the form of a beta distribution. A hypothetical beta distribution was used in a similar situation requiring estimation of variables, the PERT system. PERT (Program Evaluation and Review Technique) was developed for use by the U.S. Navy

Special Projects Department for planning and controlling the development of the Polaris submarine weapon system (Lambourne, 1967:42).

The decision to use the beta distribution in the present problem was based on the precedent set by the PERT system and the similarities between the PERT system and the present problem. The PERT system requires the user to estimate the "optimistic," "pessimistic," and "most likely" time that it will take to complete various activities; this is analogous to the estimation of "high," "low," and "most likely" figures in the present study. PERT involves the estimation of time required to achieve events together with an estimation of the uncertainties involved; the system is based on human judgment about events and times (U.S. Department of the Navy, 1958:1). Estimates for times to complete various "activities" are combined to estimate both the time needed to complete an "event" and the associated variance. The activities are analogous to the variables affecting enrollments; the events are analogous to the enrollments. The PERT model, however, avoids one of the complications of the present model: the activities are in linear combination, assuring that the

distributions of event times will be normal under the central limit theorem (Hays, 1963:242).

Assumptions were made in the PERT model about the general characteristics of the probability distribution of the time involved in performing an activity:

> It is felt that the distribution will have but one peak, and that this peak is the most likely time for completion. Thus, the point m [see Figure 3.3 (p.64)] is representative of the most probable time. Similarly, it is assumed that there is relatively little chance that either the optimistic or pessimistic estimates will be realized. Hence, small probabilities are associated with the points a and b. No assumption is made about the position of the point m relative to a and b. It is free to take any position between the two extremes -- depending entirely on the estimator's judgment [U.S. Dept. of the Navy, 1958:4] .

The Beta distribution is usually in the following form:

$$dF = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}\,dx, \quad 0 \leq x \leq 1;\ \alpha,\beta > 0$$

[Kendall and Buckland, 1957:26].

It was assumed that variables in the present model exhibit these same distribution properties. For instance, the beta distribution is unimodal. However, the distribution of a variable might not be unimodal when the variable is influenced by the probabilities of a discrete major event,

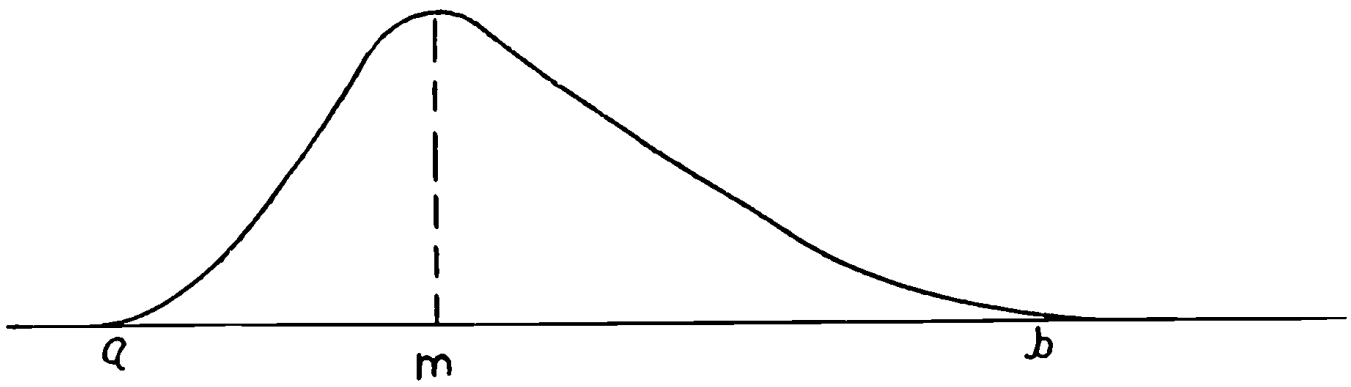

Fig. 3.3. A beta distribution.

such as the closing of a parochial school. Estimated transfers from nonpublic schools might vary around a low figure if the parochial school remained open and around a high figure if it were closed. Instead of introducing bimodal distributions into the model, it is possible to view the situation as two unimodal distributions with probabilities p and 1-p of being chosen. Thus the computer program could be modified, if necessary, to use a two step random procedure: first, drawing a distribution and second, drawing a value from the randomly chosen distribution. The beta distribution also assumes continuous values (Martin, 1968:72) and, of course, predictions of enrollment deal with discrete values. This is not considered a significant drawback, however; enrollments are considered to be continuous until the output stage of the program at which time they are truncated to integers.

Lambourne (1967) described the theoretical meaning of the optimistic, pessimistic, and most likely times:

> Suppose an individual activity were repeated under identical conditions a hundred times... and we were to plot a graph of the statistical distribution of the achieved times....We call the one shortest duration out of the hundred the Optimistic time, the one longest the Pessimistic, and the time corresponding to the high point, the Most Likely [p.43].

For the PERT model it was desirable to use as the expected values of activity times a value other than the most likely estimate, since obtained estimates showed the distribution to be skewed in many cases, usually with the likely time nearer the optimistic than the pessimistic time. An estimate of the expected value was made based on the probability density of the distribution and sample values. The estimate for the expected value of an activity time (E(t)) is as follows:

$$E(t) = (a + 4M + b)/6$$

[U.S. Department of the Navy, 1958: App.B(2), B(4)].

(See Figure 3.3 (p.64).)

For unimodal frequency distributions, the standard deviation can be estimated roughly as one-sixth of the range. Using the points a and b to represent the range, the estimate of the standard deviation becomes one-sixth of the difference between the pessimistic and optimistic time estimates:

$$SD = (b - a)/6$$

In the present use of the beta distribution, it was necessary to calculate the mean and standard deviation since each beta distribution is transformed into a normal distribution with the same mean and standard deviation. The

purpose of the transformation was to simplify the simulation process; it is much simpler to draw a random number from a normal distribution than from a beta distribution. It was assumed that a better estimate of the probability distribution is obtained if the user is allowed to describe a beta distribution which is transformed than if he is required to describe a normal distribution. The mean and standard deviation are estimated by the formulas used in the PERT model.

High and low estimates were set for the input from the Brockton study by utilizing the historical data which were given for most of the variables. In most cases, the data for the past five years were given. There was some variation in the pattern; for instance, for the female population variable, data for the past eleven years were used. The available data were used to calculate the variance across years; this variance was used to estimate the variance across hypothetical trials of the enrollment outcomes. The variance for the preschool net migration variable had to be determined arbitrarily, since adequate data were not given in the report. The variance was used to obtain nonskewed distributions by

setting a and b each three standard deviations away from m. Although a and b theoretically enclose 98 per cent of the cases, the choice of three standard deviations was made since the difference between a and b is used as an estimate of six standard deviations.

A second set of limits was chosen to represent "skewed" limits. On one side of the most likely estimate, a point three standard deviations from m was chosen; on the other side, a point five standard deviations from m was chosen. A set of skewed limits was chosen so that a kind of concurrent validity of the model could be examined. Without simulation, the "most likely" enrollment estimate would normally be used in planning; with simulation output in the form described (see Ch. I, p.8, supra.), the .50 probability figure would often be taken as the equivalent in importance for planning. To the extent that the original beta distribution of estimates is skewed, however, the .50 probability figure in the simulation and the "most likely" enrollment computed by the multivariable method without simulation may not coincide. It is assumed that the estimates in the multivariable input represent the mode. A quotation from the Center study shows that the

estimates do not seem to represent the mean or the median: "Should the recent drop in the number of building permits be suddenly reversed, this migration figure will probably be too small [Center for Field Studies, 1964: A-9]." If the migration figure were intended to be an estimate of the mean or the median, it would seem that it should be somewhat higher to take account of the possibility that the trend in building permits would be reversed. If the planner wants an estimate of the mean or median, this distortion of the most likely estimate might be desirable. However, if he actually wants an estimate of the mode, the necessary information may be lost in the transformation from the beta to the normal. The direction of the skew for each variable was chosen so that the skew would be in the direction of larger total enrollment. Thus the "errors" would not compensate each other and the full impact of the skewing would appear in the total enrollment figures.

The addition of simulation to the multivariable process provides for variation among the variables. In Figure 3.2 (p. 55), the 1963 grade 7 enrollment is adjusted to become the predicted grade 8 enrollment for 1964. For the simulation of grade 8 for 1964, net migrations, transfers to/from the

nonpublic schools, retentions, deaths, and dropouts are considered to be drawn from distributions with properties determined by the high, most likely, and low estimates. In the simulation, the multivariable method is performed with values for some or all of its variables being randomly drawn from such distributions. Since the outcome depends in part on random numbers, repetition of the process is necessary to determine statistical properties; one hundred iterations were performed for each grade and year of the simulation. After the first year of the simulation, enrollments for previous years and grades are no longer single figures, but are predictions which vary across iterations and contribute to the variance of future enrollments.

One main computer program and a subroutine for output were developed for the present study. The computer programs, MAIN and OUTPUT, were written in Fortran IV user language. MAIN also employs the subroutine GAUSS supplied by the IBM System/360 Scientific Subroutine Package (1968:77).

GAUSS computes a normally distributed random number from a distribution with a specified mean and standard deviation. Twelve uniform random numbers are used to compute normal

random numbers by the central limit theorem. The result is then adjusted to conform to the given mean and standard deviation. GUASS uses subroutine RANDU, also included in the Scientific Subroutine Package, to obtain uniform random numbers which are found by the power residue method. The random numbers generated by RANDU are actually "pseudorandom" numbers generated through a mathematical process. In computer applications, generation of pseudorandom numbers is preferable to storing a lengthy table in computer memory (Martin, 1968: 77). RANDU generates 2**29, or 536,870,912, terms before repeating the cycle (International Business Machines Corporation, 1968:77). The number of times RANDU is called in the simulation is sufficiently smaller than the number of terms in the cycle.

The first steps of MAIN give directions for reading the input data. In addition to historical data for the base years and estimates of variables for predicted years, the input data include information on the user options: the number of years ≤15 to be simulated, indication of whether or not kindergarten enrollments are to be predicted, indication of whether or not enrollments are to be calculated separately by sex, and the random number chosen to initialize the random number generator.

A print-out of the input data is produced as a check on the accuracy of the input. Next, means and standard deviations of the input variables are computed using approximation formulas (U.S. Department of the Navy, 1958: App. B(3), B(4)).

The random number generator, GAUSS, is initialized by a randomly drawn nine digit odd integer available in the input data. The initialization is optional for GAUSS, which uses its own starting value when no value is chosen by the user. In order to perform a test for reliability in the present study, it was necessary to control the starting values so that the simulation can be performed using two different randomly chosen starting values. The two starting values, 420013363 and 083632427, were drawn from a random number table (Hodgman, 1959:238-239). The program could be modified to allow a user to omit the initialization step.

Before the random number generator is used to draw values for enrollment variables, it is used to generate one thousand "throw-away" random numbers. This procedure was intended to avoid bias which might be present in the first few numbers generated.

The body of the program MAIN is enclosed in two major DO loops. The outer DO loop varies the year of simulation

from one to the number ≤ 15 specified by the user. The inner DO loop varies the grade level of the predicted enrollment from one to 12 or 13, depending on whether or not predictions for kindergarten are included.

The exact pattern of calculations for making predictions varies with the grade and year of the simulation. Within the two major DO loops, appropriate IF statements transfer control to the series of statements which calculates iterations for the particular grade and year. These series are actually DO loops which calculate the one hundred iterations for the specific grade and year; these are referred to as the four iteration DO loops. The first iteration DO loop is used in the calculations for grade K or 1 for the first five years of the simulation, or, if kindergarten enrollment is predicted, for the first four years of the simulation. The second DO loop also calculates grade K or 1 enrollment, but for the years of simulation not calculated by the first DO loop. Two slightly different methods are needed in this calculation because historical birth data are available for the first few years and must be predicted for the other years. During trial runs of the simulation, the second DO loop included a print-out of predicted births as a check on

the routine for calculating births from age-specific birth rates and numbers of women.

The third DO loop calculates enrollments for all grades above the first grade level to be predicted and for the first simulation year. The fourth DO loop calculates enrollment for these same grades for the remaining simulation years. The method of the third DO loop differs from that of the fourth loop in that the third uses historical enrollment whenever previous enrollment is required in the calculations; the fourth uses iteration-specific predicted values.

Since an iteration-specific prediction value of enrollment in one grade is used to predict the iteration-specific enrollment in the next grade and year, perhaps a more straightforward method of programming would compute predictions for all grades and years for one iteration before continuing to the next iteration. However, this would have caused complications in terms of computer memory space. All one hundred iterations for a year and grade unit must be retrieved at the same time so that their distribution may be described. However, it is not necessary to have the iterations for all grades and years stored simultaneously; only the iterations for the presently

simulated year and the previous year are necessary. The subroutine OUTPUT is called at the end of each iteration DO loop; however, provisions are made for storing iteration information for the present and the previous year.

Another instance in which information used in the prediction of one grade is stored to be used in the prediction of the next grade is the treatment of randomly drawn retention rates. Since the retention rate in grade g is used in predicting both grades g and g+1 for the following year, the randomly drawn rate is stored after use in predicting g and used again to predict g + 1. This avoids the spurious variance which would be introduced by randomly drawing another retention rate.

It should be noted that some imprecisions could be introduced in the first grade level predicted since estimates for preschool deaths are supplied to simple birth estimates rather than to birth estimates adjusted for preschool migration. Thus the preschool migration figures are not adjusted for death rates unless the preschool death rate estimate is modified to account for this; such refinement of the prediction, however, is not considered necessary.

Since the program was designed to be generalizable to various enrollment prediction situations, and since some users

may want to have output by sex, the option for breakdown by sex was included. The user has an option to give an estimated proportion male, as well as the three other estimates for each variable. The estimate for the proportion male is given as a single figure, varying only across variables; no high or low estimates are given for the estimate of the proportion in order to keep programming and input requirements as simple as possible. Of course, this introduces some bias toward underestimation of the variance for predictions.

An assumption on which the simulation was built is that of independence among the estimated variables. Random numbers from the distributions of these variables are drawn independently. In some cases, this assumption may be false, since variables such as retentions and dropouts, or births and migrations, may be related. But the development of a correction for this possibility would be more than is warranted by the exploratory nature of the present study. Variables used more than once in the computations, such as previous enrollments, are obviously not independent and are treated accordingly; the same value for the previous enrollment is used each time the enrollment is needed in the calculation.

Subroutine OUTPUT places the iterations in ascending order, computes the required percentiles, and prints the output. The fifth percentile is computed, for example, by finding the midpoint between the fifth and sixth ordered iteration. During the trial stages of the program, OUTPUT included instructions to print all one hundred iterations to assess the accuracy of the steps for ordering the iterations and computing the percentiles.

Pages 78 through 88 contain flow charts for MAIN and OUTPUT. Program steps for MAIN and OUTPUT and instructions for use of the computer programs are listed in the Appendix.

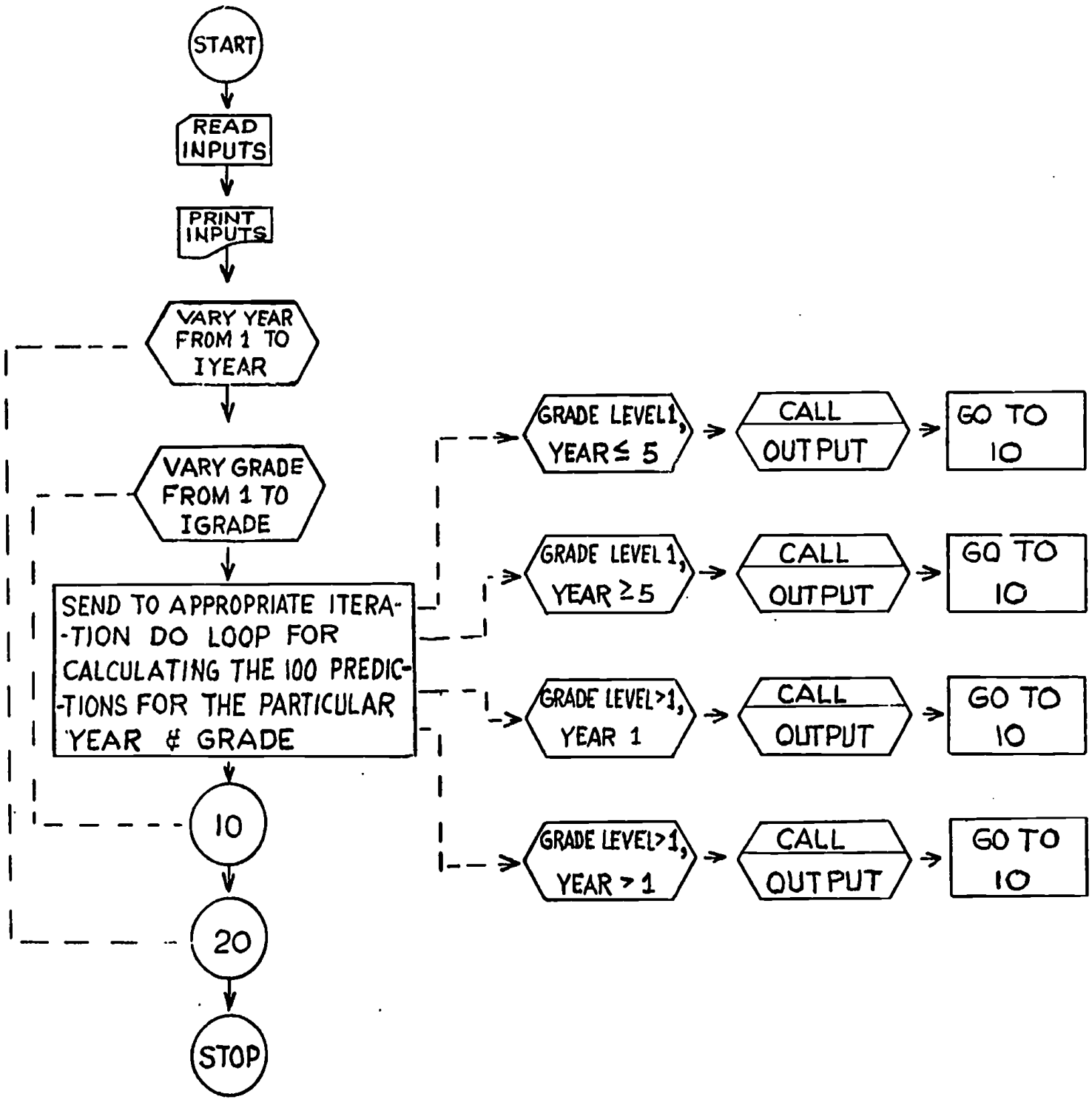

Fig. 3.4. Abridged flow chart for Program MAIN.

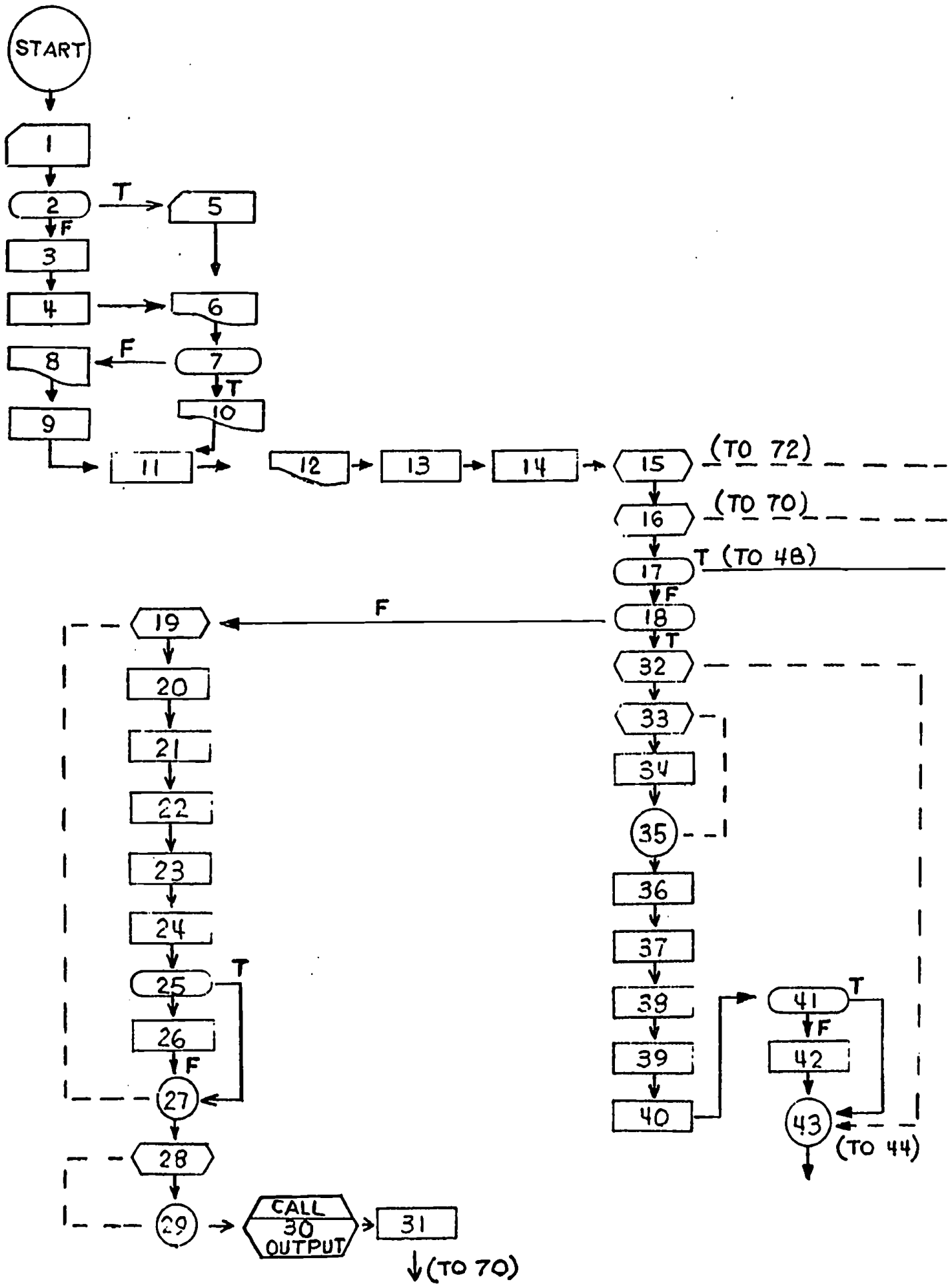

Fig. 3.5. Flow chart for Program MAIN.

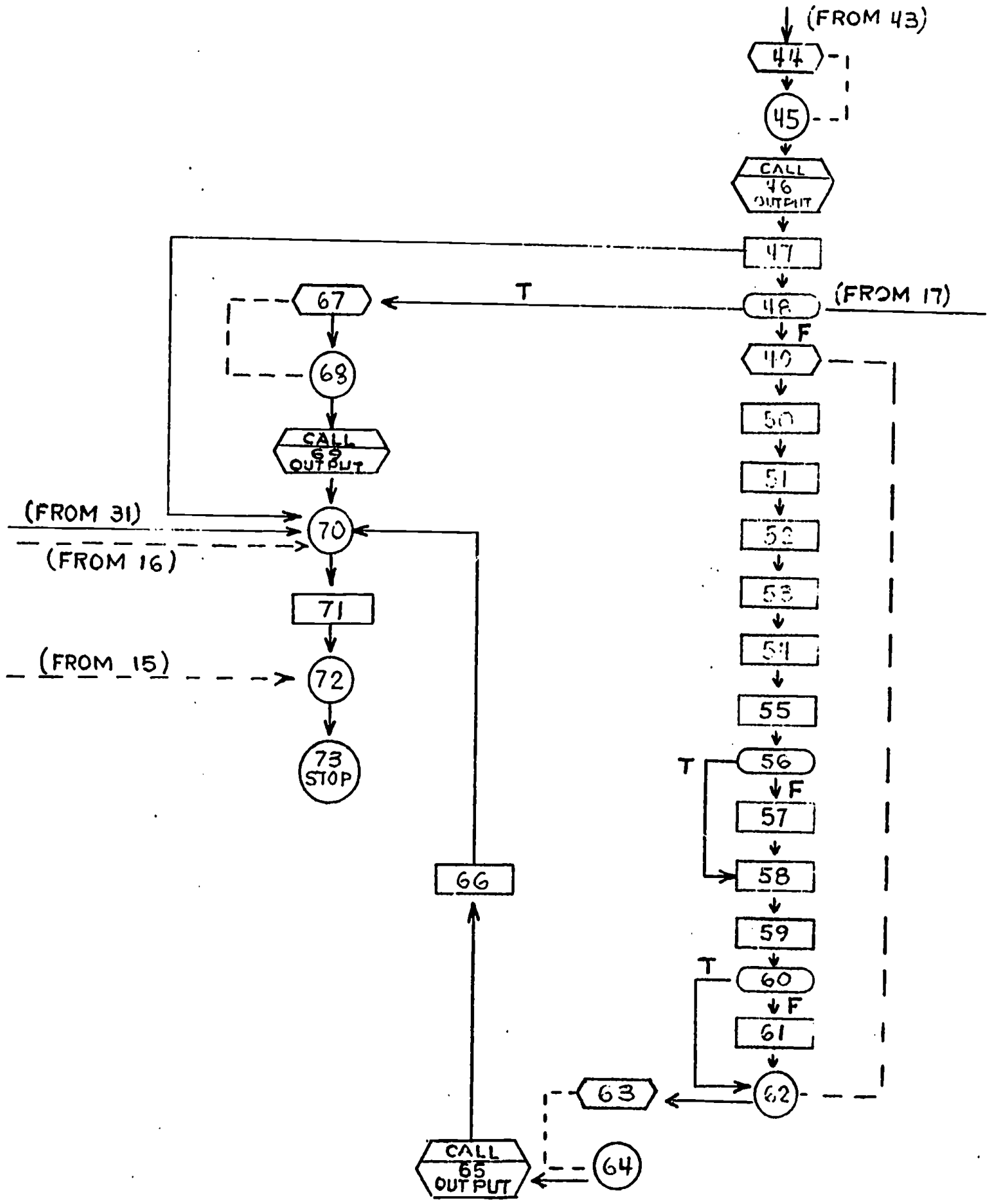

Fig. 3.5. (continued).

Figure 3.5 (continued)

Key to Flow Chart for Main Program

1. Read (a) title of job, (b) number of years to be simulated, (c) number of grades to be simulated, (d) indication of whether or not sex option is to be used, (e) beginning year of simulation, (f) integer to initialize random number generator, and (g) variable formats to be used for parameter and variable inputs.

2. If the sex option is to be used, GO TO 5.

3. Read all input parameters and variables.

4. GO TO 6.

5. Read all input parameters, including the proportion male for each input.

6. Print information which was read in step 1.

7. If sex option is used, GO TO 10.

8. Print parameter inputs (totals).

9. GO TO 11.

10. Print parameter inputs for totals and for boys.

11. Compute means and standard deviations of input variables using approximation formulas for the beta distribution.

12. Print input variables (high estimate, most likely estimate, low estimate, mean, and standard deviation). If sex option is used, print proportion boys for each variable.

13. Initialize the random number generator.

14. Generate 1000 throw-away random normal deviates.

15. Open DO loop which varies the year of simulation from 1 to IYEAR. (IYEAR is the number of years to be simulated.)

Figure 3.5 (continued)

16. Open DO loop which varies grade level from 1 to IGRADE. (IGRADE is either 12 or 13, depending on whether or not kindergarten is included. "Grade level 1" refers to either grade 1 or grade K.)

17. IF (I .GT. 1) GO TO 48. (I is the grade level.)

18. IF (J .GT. LIMIT) GO TO 32. (J is the year; LIMIT is either 4 or 5, depending on whether or not kindergarten enrollment is predicted, and it represents the number of prediction years for which historical birth data is available.)

19. Open DO loop for varying iterations from 1 to 100 in the prediction of first grade level enrollment for one of the first 4 or 5 years of simulation. This is the first of four iterations DO loops in the program.

20. Subtract deaths from births to obtain tentative prediction.

21. Add to tentative prediction a randomly drawn preschool net migration figure.

22. Subtract from tentative prediction a randomly drawn figure for nonpublic school enrollment.

23. To obtain the final prediction, add to tentative prediction the product of (a) a randomly drawn proportion of retentions in the first grade level the previous year and (b) the enrollment in the first grade level the previous year.

24. Store final prediction for iteration M, grade level, year ≤ 5.

25. If sex option is not used, GO TO 27.

26. Proceed analogously to steps 20-24 to compute predictions for boys for iteration M, grade level 1, year ≤ 5.

27. Continue. (End of first iteration DO loop, which varies iterations from 1 to 100.)

28. If sex option is not used, GO TO 30. Otherwise, open DO loop for calculating enrollment of girls for grade level 1, year ≤ 5 by subtracting male enrollment from total enrollment to obtain female enrollment.

Figure 3.5 (continued)

29. Continue. (End of DO loop for calculating female enrollment.)

30. Call subroutine OUTPUT to print probability tables for enrollment predictions for grade level 1, year ≤ 5.

31. GO TO 70.

32. Open DO loop for varying iterations from 1 to 100 in the prediction of first grade level enrollment for the remaining years of the simulation (years in which births must be predicted). This is the second iteration DO loop in the program.

33. Open DO loop for calculating births, varying age group of women from 1 to 6, by multiplying a randomly drawn birth rate by a randomly drawn number of women, and dividing by 1000. If the sex option is used, calculate male births as well as total births.

34. Accumulate the births across age groups.

35. Continue. (End of DO loop for calculating births.)

36. To obtain tentative prediction, subtract from total births the product of (a) a randomly drawn preschool death rate and (b) total births.

37. Add to tentative prediction a randomly drawn preschool net migration figure.

38. Subtract from tentative prediction a randomly drawn figure for nonpublic school enrollment.

39. To obtain final prediction, add to tentative prediction the product of (a) a randomly drawn proportion retained in the first grade level the previous year and (b) the enrollment in the first grade level the previous year.

40. Store final prediction for iteration M, grade level 1, year ≥ 5.

41. If not using sex option, GO TO 43.

42. Proceed analogously to steps 36-40 to calculate predictions for boys for the same iteration, grade, and year.

43. Continue. (End of the second iteration DO loop, which

Figure 3.5 (continued)

varies iterations from 1 to 100.)

44. If sex option is not used, GO TO 46. Otherwise, open DO loop for calculating enrollment for girls in grade level 1, year $\geq$ 5 by subtracting male enrollment from total enrollment to obtain female enrollment.

45. Continue. (End of DO loop for calculating female enrollment.)

46. Call subroutine OUTPUT to print probability tables for enrollment in grade level 1, year $\geq$5.

47. GO TO 70.

48. If (J .GT. 1) GO TO 67. (J is the year presently simulated.)

49. Open DO loop varying iterations from 1 to 100 in the prediction of enrollment for a grade level above 1 for the first year of the simulation. This is the third iteration DO loop of the program.

50. As the tenative prediction, use the historical enrollment for the previous year and grade.

51. Add to the tentative prediction a randomly drawn net migration figure.

52. Add to tentative prediction a randomly drawn figure for net transfers to public schools from nonpublic schools.

53. Subtract from tentative prediction the product of (a) the previously drawn proportion of students retained in the previous grade and year and (b) the enrollment of the previous grade and year.

54. Add to tentative prediction the product of (a) a randomly drawn proportion retained in the present grade the previous year and (b) the enrollment in the present grade the previous year.

Figure 3.5 (continued)

55. To obtain the final prediction, subtract from the tentative prediction the product of (a) a randomly drawn proportion of students dropped from rolls because of death or institutionalization during the previous year and (b) the enrollment the previous year and grade.

56. If the grade level is greater than 7, GO TO 58.

57. GO TO 59.

58. Adjust final prediction by subtracting the product of (a) a randomly drawn proportion of dropouts for the previous grade and year and (b) the enrollment in the previous grade and year.

59. Store the final predictions for iteration M, grade level >1, year 1.

60. If not using the sex option, GO TO 62.

61. Calculate the predictions for male enrollment by proceeding analogously to flow chart steps 50-59.

62. Continue. (End of the third iteration DO loop, which varies iterations from 1 to 100.)

63. If the sex option is not used, GO TO 65. Otherwise, open DO loop for calculating enrollment for girls in grade level >1, year 1 by subtracting male enrollment from total enrollment to obtain female enrollment.

64. Continue. (End of DO loop for calculating female enrollment.)

65. Call subroutine OUTPUT to print probability tables for enrollment predictions for a grade level >1, year >1.

66. GO TO 70.

67. Open DO loop for varying iterations from 1 to 100 in the prediction of enrollment for a grade level above 1 for the remaining years of the simulation. This is the fourth iteration DO loop for the program. Obtain predictions by proceeding analogously to steps 50-61, using predicted enrollment, rather than historical enrollment, for the previous year and grade.

Figure 3.5 (continued)

68. Continue. (End of fourth iteration DO loop.)

69. If using the sex option, calculate female enrollment. Then call subroutine OUTPUT.

70. Continue. (End of the DO loop which varies grade level and was opened in step 16.)

71. Store predictions across grades as "previous year" rather than "present year" predictions.

72. Continue. (End of DO loop which varied year of simulation and was opened in step 15.)

73. STOP.

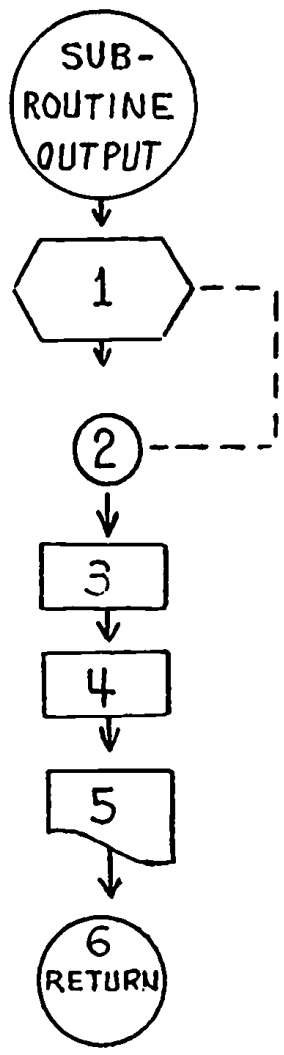

Fig. 3.6. Flow chart for Subroutine OUTPUT.

Figure 3.6 (continued)

1. Open DO loop for putting predictions in ascending order.

2. Continue. (End of ordering DO loop.)

3. Compute the required percentiles. For example, the 5th percentile is the midpoint between the 5th and 6th prediction in the ordered list.

4. Input the probability values associated with the percentiles. For example, the probability that the prediction will be less than or equal to the 5th percentile value is .05.

5. Print the percentiles and the associated probabilities.

6. RETURN to main program.

To complete the achievement of the first objective of the study, simulations were performed using enrollment and prediction data from Brockton, Massachusetts. Three factors were varied in the simulations: symmetry of the input, the random number chosen to initialize the subroutine GAUSS, and the use of the option for the prediction of enrollments separately by sex. In descriptions of the simulations, the set of input data in which the high and low estimates for the variables are equidistant from the most likely estimates is called the "symmetrical" data; the one in which they are not equidistant is called the "skewed" data. Two integers were chosen to initialize GAUSS: the first was 420013363; the second, 083632429. A simulation was performed under each at the following conditions:

(1) symmetrical data, first random number, sex option not used,

(2) skewed data, first random number, sex option not used,

(3) symmetrical data, second random number, sex option not used,

(4) symmetrical data, first random number, sex option used.

The simulations were run on the IBM System/360 Model 40 Computer at Boston College.

The second major objective was to investigate the predictive validity, reliability, and concurrent validity

of the method with the use of the Brockton enrollment data and predictions. The following four null hypotheses were formulated to test predictive validity, reliability, and concurrent validity, respectively, with the last two hypotheses outlining the two tests of concurrent validity:

(1) There is no difference between the agreement of the percentage of survival projections with the actual Brockton enrollments and the agreement of the multivariable predictions with the actual Brockton enrollments.

(2) There is no difference between the predicted enrollments produced by the two simulations using different initial values for the random number generator.

(3) There is no difference between the enrollments predicted with the multivariable method and the .50 values produced by the simulation.

(4) There is no difference between the .50 values produced by the simulation using symmetrical data and those produced using skewed data.

The research hypothesis corresponding to the first null hypothesis is that the multivariable predictions agree more closely with the actual Brockton enrollment figures than do the percentage of survival projections; this hypothesis is designed to obtain data supporting the selection of the multivariable method as the prediction method on which the

simulation was built. The second, third, and fourth research hypotheses, like the corresponding null hypotheses, are stated in terms of no differences. They are based on the idea that probable difference would indicate undesirable noise factors in the simulation output.

The predictive validity of the distributions of enrollments produced by the simulation could not be meaningfully tested since they were to a large extent determined by the high and low estimates which were chosen arbitrarily by the present investigator. The assignment of high and low estimates was necessary because data for these estimates were not available. Thus the test of predictive validity is actually a test of the predictive validity of the prediction model on which the simulation is based; it is a comparison of the predictive accuracy of the multivariable model with that of the percentage of survival model, the latter being the more frequently used method. The percentage of survival projections were computed by the present investigator from enrollment figures (Brockton School Department, 1959-1969).

Six years of historical data were used to obtain the five survival percentages which were then averaged. The number of

survival percentages to be averaged was chosen after an examination of the literature; averages of three to ten numbers are typical, and five is probably the most common (Hunt, 1967; Metcalf & Eddy, Inc., 1968).

The multivariable prediction figures were obtained by using the data from the Center for Field Studies. (1964: Appendix). Since it was necessary to make some modifications in the data for use in the simulation, the predictions calculated by the Center were not adequate for the test; the predictions were recalculated using the modified data.

To compare the accuracy of the two methods, the absolute differences between enrollments predicted by the multivariable method and actual enrollments in Brockton were compared to the absolute differences between enrollments predicted by the percentage of survival method and actual Brockton enrollments. The Wilcoxon matched-pairs signed-ranks test was chosen for this comparison. The comparison is one of two related samples of interval data. The samples were related since the predictions could be paired by grade and year. The level of measurement was

interval since the data consisted of numbers of students. There was little evidence to indicate that the data would satisfy parametric assumptions. Two nonparametric tests listed in Siegel (1956: inside covers) for data with these characteristics are the Walsh test and the randomization test for matched pairs. The Walsh test requires an assumption which does not necessarily hold for the present data. The assumption is that the differences between the matched pairs are drawn from symmetrical populations (Siegel, 1956:83). The randomization test is not based on such an assumption, but because of computational cumbersomeness, its use is recommended only for very small samples; the Wilcoxon matched-pairs signed-ranks test is suggested as an efficient alternative (Siegel, 1956:91).

Brockton enrollment figures were available for each of 12 grades for the six years 1964-1969. Comparisons were made once with an N of 72 and then separately by grade and by year with N's of six and twelve. An N of twelve, and certainly an N of 72, is large enough to make the computations cumbersome; thus the Wilcoxon test was chosen for these comparisons. It was also chosen for the comparisons with an

N of six so that the tests would be consistent and comparable.

The Wilcoxon matched-pairs signed-ranks test is actually a randomization test on the ranks requiring an ordered metric scale in which the differences between pairs can be ranked in order of absolute size (Siegel, 1956:91, 75-76). A two-tailed test was performed to avoid excluding the possibility that the percentage of survival predictions were more accurate.

The reliability was tested by comparing the two simulations whose input and options differed only in the seeds, the random numbers used to initialize GAUSS. Reliability is used in this study to refer to the relationship between the output of two runs of the simulation using two different seeds; this use of the term reliability is not to be confused with its use in some prediction studies to mean the accuracy of prediction. The reliability was tested by comparing the one hundred iterations produced for each grade and year by one simulation to those produced by the other simulation. These comparisons were made by the Kologorov-Smirnov two-sample test (Siegel, 1956:127-136). Each run of the simulation produces one hundred iterations per grade and year. With

twelve grades and twelve simulation years, the data consisted of 144 pairs of independent samples with an N of one hundred. A significance test was performed for each pair. The Kolmogorov-Smirnov test was chosen, although it assumes only ordinal data, because it is sensitive to all kinds of differences in the distributions from which the two samples are drawn; this is important since the simulation results depend on the distribution as a whole rather than just the central tendency. The Wald-Wolfowitz runs test also has this characteristic, but it probably has less power-efficiency than does the Kolmogorov-Smirnov test (Siegel, 1956:144-145). A two-tailed Kolmogorov-Smirnov test was performed using KOLM2, a program from the IBM System/360 Scientific Subroutine Package (1968:65-66).

Concurrent validity was investigated by testing null hypotheses three and four. The purpose of the third hypothesis was to detect random or systematic errors in the .50 probability level predictions resulting from the use of different pseudo-random numbers; the purpose of the fourth hypothesis was to detect error in the .50 probability predictions resulting from converting the distributions of "skewed" estimates from beta distributions to normal distributions. The Wilcoxon matched-

pairs signed-ranks test was chosen for this comparison. Like the data for the test of predictive validity, the data consisted of two related samples of interval data; the samples are related since the predictions are paired by grade and year of prediction. Other characteristics of the data relevant to choosing the test were like those of the predictive validity data. A Wilcoxon test was performed with an N of 144 (twelve grades and twelve years of prediction) using the z approximation, and tests were also performed separately by year using the T statistic. The test of hypothesis three was two-tailed, but the test of hypothesis four was one-tailed. It was possible to predict a direction of significance for hypothesis four because the input data were skewed in the direction of greater numbers of students.

Chapter IV

Results of the Study

The first objective of the study involved the collection of data, adaption of the multivariable model, writing the computer programs and instructions for the user, and performing the simulations. Discussed in Chapter III are the outcomes of all of these activities except the performance of the simulation. In Table 4.1 (p.98) and 4.2 (pp.99-101) are samples of the simulation out-out. Table 4.1 contains a sample of the output for the simulation using symmetrical data, the first random number starter, and no sex option. Table 4.2 contains a portion of the output for the simulation using symmetrical data, the first random number starter, and the sex option.

With twelve grades and twelve years of simulation, running the simulation without the sex option required 88,736 bytes in storage and 46 minutes and 13 seconds of compilation and execution time. The time and storage requirements placed the program charges at $75 per hour at the Boston College Computer Center; thus, the cost of

TABLE 4.1

PROBABILITY THAT TOTAL ENROLLMENT IN GRADE 7 IN 1971 WILL BE LESS THAN THE SPECIFIED PREDICTED ENROLLMENT

PROBABILITY	PREDICTED ENROLLMENT
.05	1512.
.10	1525.
.20	1546
.30	1559.
.40	1576.
.50	1586.
.60	1600.
.70	1614.
.80	1629.
.90	1658.
.95	1677.

PROBABILITY THAT TOTAL ENROLLMENT IN GRADE 7 IN 1971 WILL BE GREATER THAN THE SPECIFIED PREDICTED ENROLLMENT

PROBABILITY	PREDICTED ENROLLMENT
.05	1677.
.10	1658.
.20	1629.
.30	1614.
.40	1600.
.50	1586.
.60	1576.
.70	1559.
.80	1546.
.90	1525.
.95	1512.

TABLE 4.2

PROBABILITY THAT TOTAL ENROLLMENT IN GRADE 2 IN 1975 WILL BE LESS THAN THE SPECIFIED PREDICTED ENROLLMENT

PROBABILITY	PREDICTED ENROLLMENT
.05	1899.
.10	1991.
.20	2042.
.30	2103.
.40	2177.
.50	2234.
.60	2277.
.70	2351.
.80	2384.
.90	2464.
.95	2561.

PROBABILITY THAT TOTAL ENROLLMENT IN GRADE 2 IN 1975 WILL BE GREATER THAN THE SPECIFIED PREDICTED ENROLLMENT

PROBABILITY	PREDICTED ENROLLMENT
.05	2561.
.10	2464.
.20	2384.
.30	2351.
.40	2277.
.50	2234.
.60	2177.
.70	2103.
.80	2042.
.90	1991.
.95	1899.

TABLE 4.2 (continued)

PROBABILITY THAT MALE ENROLLMENT IN GRADE 2 IN 1975
WILL BE LESS THAN THE SPECIFIED PREDICTED ENROLLMENT

PROBABILITY	PREDICTED ENROLLMENT
.05	989.
.10	1031.
.20	1068.
.30	1099.
.40	1138.
.50	1165.
.60	1192.
.70	1233.
.80	1255.
.90	1304.
.95	1342.

PROBABILITY THAT MALE ENROLLMENT IN GRADE 2 in 1975
WILL BE GREATER THAN THE SPECIFIED PREDICTED ENROLLMENT

PROBABILITY	PREDICTED ENROLLMENT
.05	1342.
.10	1304.
.20	1255.
.30	1233.
.40	1192.
.50	1165.
.60	1138.
.70	1099.
.80	1068.
.90	1031.
.95	989.

TABLE 4.2 (continued)

PROBABILITY THAT FEMALE ENROLLMENT IN GRADE 2 IN 1975 WILL BE LESS THAN THE SPECIFIED PREDICTED ENROLLMENT

PROBABILITY	PREDICTED ENROLLMENT
.05	912.
.10	946.
.20	978.
.30	999.
.40	1035.
.50	1067.
.60	1095.
.70	1112.
.80	1137.
.90	1174.
.95	1204.

PROBABILITY THAT FEMALE ENROLLMENT IN GRADE 2 IN 1975 WILL BE GREATER THAN THE SPECIFIED PREDICTED ENROLLMENT

PROBABILITY	PREDICTED ENROLLMENT
.05	1204.
.10	1174.
.20	1137.
.30	1112.
.40	1095.
.50	1067.
.60	1035.
.70	999.
.80	978.
.90	946.
.95	912.

simulation without the sex option was $57.75. Simulation using the sex option required 89,224 bytes in storage and one hour 18 minutes and 54 seconds, costing $98.62.

The second major objective was the investigation of predictive validity, reliability, and concurrent validity. Statistical tests were performed for each of the four hypotheses previously stated (p. 90). The first involved a comparison of the predictive validity of the multivariable and percentage of survival methods; the second, an investigation of reliability by a comparison of the output of two different simulations. The third and fourth were set up to test concurrent validity by comparisons of the .50 values produced by the simulation with figures produced by the multivariable method without simulation and with figures produced by the simulation using skewed data.

The testing of the hypotheses resulted in statistically significant differences for all four. Although the test for predictive validity was a two-tailed test, the significant difference was in the direction of better predictive accuracy for the multivariable method, the prediction method on which the simulation is based. The finding of significant differences

for the other hypotheses, however, requires additional interpretation; since it indicates the presence of "errors," it is necessary to examine their bearing on the usefulness and validity of the simulation procedure.

The investigation of predictive validity consisted of Wilcoxon signed-ranks matched-pairs tests with the absolute differences between the actual Brockton enrollments and the multivariable predictions being compared with the absolute differences between the actual Brockton enrollments and the percentage of survival projections. Computation instructions and significance tables were found in Siegel (1956:75-83, 254). Siegel recommended that a z approximation to the T, the usual statistic for the Wilcoxon, be used for sample sizes over 25. The level of significance chosen for rejection of the null hypothesis was .05. Using a two-tailed test and an N of 72, combining the 12 grades and six years of prediction, the value of z was -2.29, which is significant at the .022 probability level, indicating better predictive accuracy for the multivariable method. Analyzing the data separately by year of prediction, there were six significance tests with an N of 12. Although the values of T for five of

six tests were in the direction of better predictive validity for the multivariable method, none of the T values were significant. When the data were analyzed separately by grade level with N's of six, the values of T for grades 2, 3, 4, and 5 were significant at or beyond the .05 level in the direction of better predictive validity for the multivariable method. The values of T for the other grades were not significant; five were in the direction of better predictive validity for the multivariable method; grades 7, 8, and 12 were in the opposite direction. Since the test using an N of 72 was significant at the .022 level, the hypothesis of no difference in predictive validity was rejected.

In summary, the significance of the N of 72 can be viewed as composed of the results analyzed separately by year, which are for the most part in the right direction, but which lack a large enough N for significance. It can also be viewed as a summary of the significance tests computed separately by grade. Further studies are needed to interpret the patterns of significance, lack of significance, and direction of the differences in the significance

tests for the separate years and grades. For example, as more years of enrollment figures are available for Brockton, the tests computed separately by grade can be conducted with a larger N. Examination of the determinants of the Brockton enrollments might show why in some instances the percentage of survival method was a better predictor than the multivariable method.

The significance test used for testing the reliability was the Kolmogorov-Smirnov two-sample test, which measures agreement between two cumulative distributions (Siegel, 1956:127-136). In the present case, the two distributions were the 100 predictions produced by the simulation for a given grade and year using one random number starter and the 100 predictions produced for the same grade and year using a different random number starter. Since the simulation produced output for 12 grades and 12 years, 144 tests of significance were performed. The level of significance chosen for rejection of the null hypothesis was .05. Only 7 of the 144 failed to reach significance at the .05 level or beyond. Thus the null hypothesis of no difference in outputs of the simulation was rejected. The computer program printed the significance levels correct to five decimal places;

significance levels for 109 of the 144 tests were .00000.

The significance test used for the third and fourth hypotheses was the Wilcoxon matched-pairs signed ranks test. Since the Kolmogorov-Smirnov test showed that the simulation outputs using different pseudo-random numbers were significantly different, the Wilcoxon tests for the third hypothesis were performed separately on the outputs of the simulations differing in the pseudo-random numbers used. Two-tailed tests were performed; as in the tests of all four hypotheses, the .05 level of significance was chosen as the rejection level. The T statistic was used for the tests performed separately by year; the z approximation was used for the overall tests of significance. The tests with the data from the simulation using the first random number showed a z of -3.832, which is significant beyond the .05 level; however, none of the T's were significant at .05 although all were in the direction of higher predictive figures for the simulation data. The output using the second random number produced a z of -2.7575, also significant beyond the .05 level. Only one year, 1966, showed significant differences at the $\leq .05$ level; for all the years except 1972

and 1973 the differences were in the direction of higher predictive figures for simulation data. Since the overall significance tests for both sets of data produced significant z's, the third null hypothesis was rejected, indicating a lack of perfect correspondence between the 0.50 simulation figures and the nonsimulation figures.

The Wilcoxon tests for the fourth null hypothesis, testing the effect of skewing the input, were one-tailed tests, since the directional effect of the skewing could be predicted. As in the tests of the third null hypothesis, the T statistic was used for tests performed separately by year; the z approximation was used for overall tests of significance. The tests produced a z of -10.41; the z and all of the T's were significant beyond the 0.05 level. These results indicated rejection of the fourth null hypothesis.

Interpretation of the significance tests for the null hypotheses is aided by considering their impact on the simulation output in terms of numbers and percentages of students. Tables 4.3, 4.4, and 4.5 (pp.109-121) contain the data used in the tests of predictive validity, reliability, and concurrent validity, respectively. Table 4.6

(p. 122) contains a summary of the mean algebraic and arithmetic differences between columns of enrollment figures on the three preceding tables and includes the corresponding percentages. Table 4.6 is useful for comparing the impact on the simulation results of the significant differences found by the tests of significance.

The percentage of survival and multivariable model produced significantly different predictions. However, the extent of the differences can be examined in other ways. One measure of prediction accuracy was defined by Greenawalt and Mitchell: "It was assumed that a forecast, having run seven years, which predicted enrollment within plus or minus 10 per cent of the actual enrollment was 'accurate' [1966:8]." Using the ten percent standard, nine of the 72 predictions using the multivariable method were "inaccurate"; ten of the 72 percentage of survival projections were "inaccurate." The figures which were not within ten percent of the actual enrollments are marked by asterisks in Table 4.3. The scoring procedure used here was somewhat different from that of Greenwalt and Mitchell, since the prediction period was six, rather than seven, years and accuracy scores were calculated for all grades and years,

TABLE 4.3

The Comparison of the Actual Brockton Enrollments
With Those Predicted by the Multivariable Technique
and Those Projected by the Percentage of Survival Technique

A = Actual Enrollments in Brockton Public Schools

B = Enrollments Predicted by the Multivariable Technique Without Simulation

C = Enrollments Projected by the Percentage of Survival Technique

Date 10/1/64

Grade	A	B	C	A-B	A-C	|A-B| - |A-C|
1	1639	1662	1649	- 23	- 10	+ 13
2	1488	1521	1531	- 33	- 43	- 10
3	1391	1423	1434	- 32	- 43	- 11
4	1306	1324	1329	- 18	- 23	- 5
5	1240	1246	1278	- 6	- 38	- 32
6	1157	1192	1211	- 35	- 54	- 19
7	1131	1170	1214	- 39	- 83	- 44
8	1276	1175	1172	+ 1	+ 4	- 3
9	1064	1084	1064	- 20	0	+ 20
10	1055	927	992	-128*	+ 63	+ 65
11	1052	1081	1042	- 29	+ 10	+ 19
12	931	990	924	- 59	+ 7	+ 52
		Arithmetic mean =		35.25	31.50	24.42
		Algebraic mean =		-13.75	-17.50	3.75

Date 10/1/65

1	1659	1730	1757	- 71	- 98	- 27
2	1580	1559	1555	+ 21	+ 25	- 4
3	1475	1519	1548	- 44	- 73	- 29
4	1375	1408	1521	- 33	- 46	- 13
5	1300	1311	1351	- 11	- 51	- 40
6	1233	1235	1288	- 2	- 55	- 53
7	1195	1207	1275	- 12	- 80	- 68
8	1260	1172	1211	+ 88	+ 49	+ 39
9	1130	1145	1131	- 15	- 1	+ 14
10	1055	996	1050	+ 59	+ 5	+ 54
11	981	888	915	+ 93	+ 66	+ 27
12	882	1022	922	-140*	- 40	+100
		Arithmetic mean =		49.08	49.08	39.00
		Algebraic mean =		- 5.58	-24.92	0.00

TABLE 4.3 (continued)

Date 10/1/66

Grade	A	B	C	A-B	A-C	|A-B| - |A-C|
1	1672	1836	1871	-164	-199*	- 35
2	1596	1622	1657	- 26	- 61	- 35
3	1564	1560	1572	+ 4	- 8	- 4
4	1452	1503	1534	- 51	- 82	- 31
5	1367	1395	1445	- 28	- 78	- 50
6	1332	1300	1362	+ 32	- 30	+ 2
7	1276	1251	1356	+ 25	- 80	- 55
8	1208	1208	1272	0	- 64	- 64
9	1097	1173	1168	- 76	- 71	+ 5
10	1127	952	1116	+175*	+ 11	+164
11	1019	924	969	+ 95	+ 50	+ 45
12	829	817	810	+ 12	+ 19	- 7
	Arithmetic mean =			57.33	62.75	41.42
	Algebraic mean =			- 0.17	-49.42	- 5.42

Date 10/1/67

1	1804	1976	2009	-172	-205*	- 35
2	1612	1719	1764	-107	-152	- 45
3	1567	1622	1675	- 55	-108	- 53
4	1468	1545	1558	- 77	- 90	- 13
5	1452	1488	1560	- 36	-108	- 72
6	1435	1382	1457	+ 53	- 22	+ 31
7	1398	1316	1434	+ 82	- 36	+ 46
8	1312	1252	1353	+ 60	- 41	+ 19
9	1095	1210	1227	-115*	-132*	- 17
10	1013	1002	1153	+ 11	-140*	-129
11	1013	939	1029	+ 74	- 16	+ 58
12	866	889	857	- 23	+ 9	+ 14
	Arithmetic mean =			72.08	88.25	44.17
	Algebraic mean =			-25.41	-86.75	-16.17

Date 10/1/68

1	1718	1990	2009	-272*	-291*	- 19
2	1725	1850	1894	-125	-169	- 44
3	1643	1718	1783	- 75	-140	- 65
4	1563	1606	1660	- 43	- 97	- 54
5	1519	1531	1584	- 12	- 65	- 53
6	1460	1475	1573	- 15	-113	- 98
7	1590	1399	1534	+191*	+ 56	+135
8	1404	1316	1431	+ 88	- 27	+ 61
9	1158	1249	1305	- 91	-147*	- 56
10	1041	1045	1211	- 4	-170*	-166
11	935	992	1064	- 57	-129*	- 72
12	928	907	910	+ 21	+ 18	+ 3
	Arithmetic mean =			82.33	118.50	68.83
	Algebraic mean =			-32.83	-106.17	-36.67

TABLE 4.3 (continued)

Date 10/1/69

Grade	A	B	C	A-B	A-C	|A-B| - |A-C|
1	1828	2049	2069	-221*	-241*	- 20
2	1731	1867	1894	-136	-163	- 27
3	1784	1847	1915	- 63	-131	- 68
4	1620	1699	1767	- 79	-147	- 68
5	1596	1590	1688	+ 6	- 92	- 86
6	1543	1517	1597	+ 26	- 54	- 28
7	1703	1492	1656	+211*	+ 47	+164
8	1570	1399	1530	+171*	+ 40	+131
9	1329	1324	1380	+ 5	- 51	- 46
10	1147	1057	1288	+ 91	-141*	- 50
11	1033	1033	1117	0	- 84	- 84
12	919	963	941	- 44	- 22	+ 22
		Arithmetic mean =		87.75	101.08	71.17
		Algebraic mean =		- 2.75	- 86.58	- 8.33

TABLE 4.4

The 0.50 Probability Figures Calculated by Simulation With Symmetrical Data Using Different Seeds for Random Number Generator

Date 10/1/64

Grade	Random Seed #1	Random Seed #2	Random Seed #1 -Random Seed #2
1	1668	1670	- 2
2	1520	1519	+ 1
3	1415	1419	- 4
4	1332	1332	0
5	1247	1245	+ 2
6	1190	1190	0
7	1172	1170	+ 2
8	1174	1174	0
9	1076	1087	-11
10	933	929	+ 4
11	1082	1083	- 1
12	993	992	+ 1
	Arithmetic mean =		2.33
	Algebraic mean =		-0.67

Date 10/1/65

1	1728	1744	-16
2	1562	1565	- 3
3	1519	1520	- 1
4	1403	1400	+ 3
5	1319	1318	+ 1
6	1235	1234	+ 1
7	1204	1206	- 2
8	1174	1173	+ 1
9	1149	1142	+ 7
10	986	1012	-26
11	894	887	+ 7
12	1027	1025	+ 2
	Arithmetic mean =		5.83
	Algebraic mean =		-2.17

TABLE 4.4 (continued)

Date 10/1/66

Grade	Random Seed #1	Random Seed #2	Random Seed #1 -Random Seed #2
1	1845	1839	+ 6
2	1620	1632	-12
3	1561	1566	- 5
4	1503	1501	+ 2
5	1391	1392	- 1
6	1309	1306	+ 3
7	1251	1251	0
8	1206	1206	0
9	1178	1179	- 1
10	952	961	- 9
11	921	939	-18
12	822	815	+ 7
	Arithmetic mean =		5.33
	Algebraic mean =		-2.33

Date 10/1/67

1	1997	1983	+14
2	1726	1724	+ 2
3	1615	1631	-16
4	1556	1552	+ 4
5	1490	1488	+ 2
6	1376	1378	- 2
7	1321	1323	- 2
8	1251	1250	+ 1
9	1192	1199	- 7
10	1011	1009	+ 2
11	954	957	- 3
12	882	904	-22
	Arithmetic mean =		6.42
	Algebraic mean =		-2.25

Date 10/1/68

1	1993	1994	- 1
2	1865	1856	+ 9
3	1718	1727	- 9
4	1595	1617	-22
5	1542	1550	- 8
6	1478	1476	+ 2
7	1396	1398	- 2
8	1323	1320	+ 3
9	1258	1254	+ 4
10	1041	1030	+11
11	1006	991	+15
12	914	925	-11
	Arithmetic mean =		8.08
	Algebraic mean =		-0.75

TABLE 4.4 (continued)

Date 10/1/69

Grade	Random Seed #1	Random Seed #2	Random Seed #1 -Random Seed #2
1	2061	2049	+12
2	1869	1874	- 5
3	1863	1840	+23
4	1705	1712	- 7
5	1583	1605	-22
6	1527	1530	- 3
7	1491	1494	- 3
8	1397	1396	+ 1
9	1318	1325	- 7
10	1092	1093	- 1
11	1020	1023	- 3
12	967	954	+13
	Arithmetic mean =		8.33
	Algebraic mean =		-0.17

Date 10/1/70

1	2099	2033	+66
2	1923	1922	+ 1
3	1865	1873	- 8
4	1843	1814	+29
5	1685	1701	-16
6	1566	1589	-23
7	1545	1555	-10
8	1491	1491	0
9	1471	1455	+16
10	1139	1153	-14
11	1072	1086	-14
12	982	990	- 8
	Arithmetic mean =		17.08
	Algebraic mean =		1.58

Date 10/1/71

1	2139	2149	-10
2	1988	1920	+68
3	1920	1924	- 4
4	1844	1856	-12
5	1824	1802	+22
6	1669	1684	-15
7	1586	1607	-21
8	1544	1551	- 7
9	1587	1576	+11
10	1290	1281	+ 9
11	1122	1144	-22
12	1038	1050	-12
	Arithmetic mean =		17.75
	Algebraic mean =		0.58

TABLE 4.4 (continued)

Date 10/1/72

Grade	Random Seed #1	Random Seed #2	Random Seed #1 -Random Seed #2
1	2228	2207	+21
2	1997	2011	-14
3	1974	1925	+49
4	1915	1898	+17
5	1830	1843	-13
6	1807	1788	+19
7	1688	1702	-14
8	1583	1610	-27
9	1645	1653	- 8
10	1400	1390	+10
11	1273	1265	+ 8
12	1083	1095	-12
	Arithmetic mean =		17.67
	Algebraic mean =		3.00

Date 10/1/73

1	2286	2287	- 1
2	2083	2051	+32
3	1997	2000	- 3
4	1956	1908	+48
5	1907	1888	+19
6	1818	1824	- 6
7	1829	1809	+20
8	1686	1696	-10
9	1678	1707	-29
10	1447	1459	-12
11	1379	1367	+12
12	1224	1224	0
	Arithmetic mean =		16.00
	Algebraic mean =		5.83

Date 10/1/74

1	2384	2393	- 9
2	2149	2144	+ 5
3	2087	2052	+35
4	1970	1976	- 6
5	1937	1891	+46
6	1894	1872	+22
7	1840	1849	- 9
8	1827	1812	+15
9	1791	1794	- 3
10	1490	1521	-31
11	1428	1433	- 5
12	1321	1317	+ 4
	Arithmetic mean =		15.83
	Algebraic mean =		5.33

TABLE 4.4 (continued)

Date 10/1/75

Grade	Random Seed #1	Random Seed #2	Random Seed #1 -Random Seed #2
1	2457	2412	+45
2	2234	2230	+ 4
3	2151	2147	+ 4
4	2070	2038	+32
5	1949	1950	- 1
6	1917	1883	+34
7	1917	1897	+20
8	1842	1850	- 8
9	1902	1908	- 6
10	1580	1593	-13
11	1465	1494	-29
12	1373	1376	- 3
	Arithmetic mean =		16.58
	Algebraic mean =		6.58

TABLE: 4.5

The Comparison of the Non-simulation Multivariable Predictions with the 0.50 Probability Figures for the Symmetrical and the Skewed Data

A = Enrollments Predicted by Multivariable Technique Without Simulation

B = 0.50 Probability Level Prediction Calculated by Simulation Using Symmetrical Data

C = 0.50 Probability Level Prediction Calculated by Simulation Using Skewed Data

Date 10/1/64

Grade	A	B	C	A-B	B-C
1	1662	1668	1695	- 6	- 27
2	1521	1520	1521	+ 1	- 1
3	1423	1415	1417	+ 8	- 2
4	1324	1332	1334	- 8	- 2
5	1246	1247	1249	- 1	- 2
6	1192	1190	1191	+ 2	- 1
7	1170	1172	1174	- 2	- 2
8	1175	1174	1175	+ 1	- 1
9	1084	1076	1091	+ 8	- 15
10	927	933	947	- 6	- 14
11	1081	1082	1095	- 1	- 13
12	990	993	1001	- 3	- 8
	Arithmetic mean =			3.92	7.33
	Algebraic mean =			-0.58	-7.33

Date 10/1/65

1	1730	1728	1757	+ 2	- 29
2	1559	1562	1592	- 3	- 30
3	1519	1519	1521	0	- 2
4	1408	1403	1406	+ 5	- 3
5	1311	1319	1323	- 8	- 4
6	1235	1235	1240	0	- 5
7	1207	1204	1206	+ 3	- 2
8	1172	1174	1177	- 2	- 3
9	1145	1149	1165	- 4	- 16
10	996	986	1009	+ 10	- 25
11	888	894	919	- 6	- 25
12	1022	1027	1048	- 5	- 21
	Arithmetic mean =			3.33	13.75
	Algebraic mean =			0	-13.75

TABLE 4.5 (continued)

Date 10/1/66

Grade	A	B	C	A-B	B-C
1	1836	1845	1879	- 9	- 34
2	1622	1620	1647	+ 2	- 27
3	1560	1561	1590	- 1	- 29
4	1503	1503	1509	0	- 6
5	1395	1391	1394	+ 4	- 3
6	1300	1309	1314	- 9	- 5
7	1251	1251	1257	0	- 6
8	1208	1206	1210	+ 2	- 4
9	1173	1178	1199	- 5	- 21
10	952	952	983	0	- 31
11	924	921	958	+ 3	- 37
12	817	822	853	- 5	- 31
	Arithmetic mean =			3.33	19.50
	Algebraic mean =			-1.5	-19.50

Date 10/1/67

1	1976	1997	2032	- 21	- 35
2	1719	1726	1757	- 7	- 31
3	1622	1615	1644	+ 7	- 29
4	1545	1556	1584	- 11	- 28
5	1488	1490	1494	- 2	- 4
6	1382	1376	1382	+ 6	- 6
7	1316	1321	1328	- 5	- 7
8	1251	1251	1258	+ 1	- 7
9	1210	1192	1209	+ 18	- 17
10	1002	1011	1045	- 9	- 34
11	939	954	999	- 15	- 45
12	889	882	923	+ 7	- 41
	Arithmetic mean =			9.08	23.67
	Algebraic mean =			-2.58	-23.67

Date 10/1/68

1	1990	1993	2026	- 3	- 33
2	1850	1856	1902	- 6	- 46
3	1718	1718	1756	0	- 38
4	1606	1595	1625	+ 11	- 30
5	1531	1542	1575	- 11	- 33
6	1475	1478	1483	- 3	- 5
7	1399	1396	1403	+ 3	- 7
8	1316	1323	1331	- 7	- 8
9	1249	1258	1285	- 9	- 27
10	1045	1041	1073	+ 4	- 32
11	992	1006	1053	- 14	- 47
12	907	914	963	- 7	- 49
	Arithmetic mean =			6.50	32.08
	Algebraic mean =			-3.50	-32.08

TABLE 4.5 (continued)

Date 10/1/69

Grade	A	B	C	A-B	B-C
1	2049	2061	2285	- 12	-224
2	1867	1869	1898	- 2	- 29
3	1847	1863	1897	- 16	- 34
4	1699	1705	1739	- 6	- 34
5	1590	1583	1616	- 7	- 33
6	1517	1527	1561	- 10	- 34
7	1492	1491	1499	+ 1	- 8
8	1399	1397	1404	+ 2	- 7
9	1324	1318	1344	+ 6	- 26
10	1056	1092	1130	- 36	- 38
11	1033	1020	1058	+ 13	- 38
12	963	967	1018	- 4	- 51
	Arithmetic mean =			9.58	46.33
	Algebraic mean =			-5.92	-46.33

Date 10/1/70

1	2090	2099	2338	- 9	-239
2	1922	1923	2119	- 1	-196
3	1869	1865	1895	+ 4	- 30
4	1827	1843	1880	- 16	- 37
5	1683	1685	1715	- 2	- 30
6	1576	1566	1601	+ 10	- 35
7	1537	1545	1581	- 8	- 36
8	1491	1491	1502	0	- 11
9	1471	1471	1504	0	- 33
10	1151	1139	1174	+ 12	- 35
11	1045	1072	1119	- 27	- 47
12	997	982	1023	+ 15	- 41
	Arithmetic mean =			8.67	64.17
	Algebraic mean =			-1.83	-64.17

Date 10/1/71

1	2151	2139	2380	+ 13	-241
2	1960	1988	2218	- 28	-230
3	1922	1920	2111	+ 2	-191
4	1851	1844	1875	+ 7	- 31
5	1808	1824	1861	- 16	- 37
6	1667	1669	1702	- 2	- 33
7	1595	1586	1621	+ 9	- 35
8	1538	1544	1578	- 6	- 34
9	1585	1587	1617	- 2	- 30
10	1288	1290	1330	- 2	- 40
11	1136	1122	1163	+ 14	- 41
12	1009	1038	1088	- 29	- 50
	Arithmetic mean =			10.83	82.75
	Algebraic mean =			-3.33	-82.75

TABLE 4.5 (continued)

Date 10/1/72

Grade	A	B	C	A-B	B-C
1	2220	2228	2460	- 8	-232
2	2018	1997	2224	+ 21	-227
3	1961	1974	2203	- 13	-229
4	1904	1915	2099	- 11	-184
5	1835	1830	1865	+ 5	- 35
6	1792	1807	1847	- 15	- 40
7	1687	1688	1726	- 1	- 38
8	1595	1583	1620	+ 12	- 37
9	1640	1645	1696	- 5	- 51
10	1393	1400	1432	- 7	- 32
11	1268	1273	1325	- 5	- 52
12	1096	1083	1129	+ 13	- 46
	Arithmetic mean =			9.67	100.25
	Algebraic mean =			-1.17	-100.25

Date 10/1/73

1	2289	2286	2532	+ 3	-246
2	2081	2083	2308	- 2	-225
3	2018	1997	2226	+ 21	-229
4	1942	1956	2185	- 14	-229
5	1886	1907	2094	- 21	-187
6	1818	1818	1852	0	- 34
7	1813	1829	1873	- 16	- 44
8	1686	1686	1724	0	- 38
9	1695	1678	1736	+ 17	- 58
10	1443	1447	1517	+ 4	- 70
11	1368	1379	1422	- 11	- 43
12	1220	1224	1281	- 4	- 57
	Arithmetic mean =			9.42	121.67
	Algebraic mean =			-2.58	-121.67

Date 1-/1/74

1	2359	2384	2643	- 25	-259
2	2146	2149	2377	- 3	-228
3	2081	2087	2320	- 6	-233
4	1999	1970	2193	+ 29	-223
5	1924	1937	2162	- 13	-225
6	1869	1894	2067	- 25	-173
7	1842	1840	1875	+ 2	- 35
8	1811	1827	1867	- 16	- 40
9	1782	1791	1848	- 9	- 57
10	1493	1490	1551	+ 3	- 61
11	1417	1428	1499	- 11	- 71
12	1315	1321	1374	- 6	- 53
	Arithmetic mean =			12.33	138.17
	Algebraic mean =			- 6.67	-138.17

TABLE 4.5 (continued)

Date 10/1/75

Grade	A	B	C	A-B	B-C
1	2428	2457	2719	- 29	-262
2	2211	2234	2468	- 23	-234
3	2145	2151	2388	- 6	-237
4	2060	2070	2308	- 10	-238
5	1980	1949	2171	+ 31	-222
6	1906	1917	2142	- 11	-225
7	1893	1917	2096	- 24	-179
8	1825	1842	1881	- 17	- 39
9	1904	1902	1964	+ 2	- 62
10	1572	1580	1656	- 8	- 76
11	1466	1465	1541	+ 1	- 76
12	1362	1373	1445	- 11	- 72
	Arithmetic mean =			14.42	160.17
	Algebraic mean =			- 8.75	-160.17

TABLE 4.6

Means and Corresponding Percentages Computed for Data Used in the Tests of the Four Null Hypotheses

Hypothesis # 1

	Arithmetic Mean	% of Actual Enrollment	Algebraic Mean	% of Actual Enrollment
Actual enrollments minus Multivariable Predictions (A-B)	63.97	4.8	-13.42	1.0
Actual enrollments minus Percentage of Survival Projections (A-C)	75.19	5.6	-61.89	4.6
|A-B| - |A-C|	48.17	3.6	-10.47	.80

Hypothesis # 2

	Arithmetic Mean	% of Seed #1	Algebraic Mean	% of Seed #1
Seed #1 minus Seed #2 (1st 6 years)	6.05	.44	- 1.39	.10
Seed #1 minus Seed #2 (all years)	11.44	.74	1.21	.08

Hypothesis # 3

	Arithmetic Mean	% of Seed #1 (Symmetrical)	Algebraic Mean	% of Seed #1 (Symmetrical)
Multivariable minus Symmetrical (1st 6 years)	5.96	.44	- 2.35	.17
Multivariable minus Symmetrical (all years)	8.42	.54	- 3.20	.21

Hypothesis # 4

	Arithmetic Mean	% of Seed #1 (Symmetrical)	Algebraic Mean	% of Seed #1 (Symmetrical)
Symmetrical minus Skewed (1st 6 years)	23.78	1.77	-23.78	1.77
Symmetrical minus Skewed (all years)	67.49	4.38	-67.49	4.38

rather than for the total enrollment for the last prediction year. However, the figures indicated that the differences in the accuracy of the two methods were not large when using this definition of accuracy.

Perhaps the most meaningful way to interpret the finding of significant difference is to compare the differences in terms of numbers and percentages of students found under this hypothesis to those found under the other three hypotheses. (See Table 4.6.) Since predictive validity data were available for only six years, the data were compared to those for the first six years for the other tests. Absolute differences between the predicted and actual enrollment figures were calculated for both the multivariable and percentage of survival methods (Table 4.3). Algebraic differences between these absolute differences are also reported in Table. 4.3. The arithmetic mean of this column is 48.17; the algebraic mean is -10.47. These figures represent the variation in the output attributable to the use of different prediction models; the variation is large compared to that which can be attributed to the use of different pseudo-random numbers, where the mean arithmetic difference is 6.05 and the mean algebraic difference is -1.39. It also makes a large difference when compared to

errors introduced by using simulation instead of direct calculation of the multivariable method; the mean arithmetic difference for this factor for the first six years is 5.96, and the mean algebraic difference is -2.35. However, the variation in output caused by skewing the input data is somewhat larger; the mean arithmetic difference is 23.78; the mean algebraic difference is -23.78. If algebraic means are compared, the impact of concurrent validity as measured by comparing skewed and symmetrical predictions was greater than that made by the choice of the prediction model. Algebraic means, instead of arithmetic means, were compared since the algebraic mean best represents the advantage of one prediction method over the other. One can say that the choice of the prediction model made a difference in the output of predictions over and above the noise caused by imperfect reliability and the use of simulation, but the effect was somewhat less than the effect caused by skewed data.

Only 7 of the 144 Kolmogorov-Smirnov tests of significance for reliability failed to reach significance at the 0.05 level or beyond. The only difference in the program or the input was the choice of the random number to initialize the GAUSS, the random number generator. The distributions produced by the two simulations should be identical if the pseudo-random numbers are truly random and the number of iterations used in the simulation is large enough. The lack

of correspondence between the two sets of distributions raised several questions:

(1) Were the initial random numbers chosen properly?

(2) Is the random number generator functioning properly according to other tests of randomness?

(3) To what extent is the output of the simulation affected by the imperfect reliability?

(4) What steps could be taken to improve the reliability?

The investigator checked to make sure that the specifications for choosing the initial random numbers were followed exactly. GAUSS requires an odd integer with nine or less digits (International Business Machines Corporation, 1968:77 and 1959:5).

In order to answer the question about the proper functioning of the random number generator, pseudo-random numbers were generated using each of the two seeds. The numbers generated using each seed were considered to be scores on 20 variables for 500 people; scores were generated within persons. Intercorrelations among the variables were computed for both sets of data. The two sets of means and

variances of the variables were compared by t tests and F tests. The matrices were compared by z tests for correlations.

Twenty t tests were performed to compare the means of the twenty variables. Only one was significant; the value of t was 2.71, significant at $\leq .01$. Of 20 F tests comparing the variances of the variables in one matrix to those of the other matrix, 4 were significant at $\leq .05$, indicating greater variance among the numbers produced using the first seed. Intercorrelations of variables within the matrices were compared across matrices. Using z tests for the difference between correlations, only 2 of the 190 tests were significant at $\leq .05$. Thus only the F tests produced much evidence that the pseudo-random numbers might not be adequately random. However, there were some differences in the numbers used in those tests and the numbers used in the simulation. The tests used only 10,000 pseudo-random numbers produced by each seed; the simulation employed approximately 100,000. The simulation also used the technique of generating 1000 throw-away numbers before generating the numbers to use in the predictions. More significant is the fact that these tests of randomness are only a few of the many tests which could have been performed

(International Business Machines Corporation, 1959:7-8).

Perhaps the most meaningful way to examine the problem of reliability is to determine the extent to which the output of the simulation is affected by imperfect reliability. One measure of this is the comparison of the fit of the distributions to the actual Brockton enrollment figures. For this measure the actual enrollment figures were compared to the eleven specified probability levels of the predicted enrollments. An actual enrollment figure was considered to have the same probability in both distributions if it fell either on or between the same probability levels; using his standard, sixty per cent of the actual enrollment figures had the same probability. An enrollment figure was considered to have nearly the same probability in both distributions if it satisfied the above standard or if it fell exactly on a specified probability level in one distribution and in the interval next to it in the other; 79.2 per cent of the enrollments satisfied this criterion.

Table 4.6 expresses the lack of reliability in terms of numbers and percentages of students. For both the calculations for the first six years and those for all twelve years, the arithmetic and algebraic means were less than those computed

to determine the effects of the prediction model and of skewing the input data. Percentages of students involved in the reliability calculations ranged from 0.08 per cent, corresponding to the algebraic mean for all twelve years, to 0.74 per cent, corresponding to the arithmetic mean for all twelve years. Percentage figures for calculations determining the effects of the prediction model and skewing ranged from 0.80 per cent, to 5.6 per cent. Thus the lack of perfect reliability made a minimal difference in the outputs relative to the differences made by the choice of model and skewing of the input. The percentages obtained from the reliability calculations were approximately the same as those obtained from the calculations involving the differences between simulated and non-simulated output. (See Table 4.6) In fact, all of these percentages seem small when compared to the standard of accuracy of enrollment prediction used by Greenawalt and Mitchell (1966:8): that enrollment predictions within plus or minus ten per cent of the actual enrollment are considered accurate.

There are several approaches that could be tried to increase the reliability or to diminish the effect of unreliability on the model. One would be to try a different random number generator; another would be to increase the

number of iterations in the simulation so that random drawings from distributions of variables would better represent the distributions. Also, the format of the output could be modified to make it less sensitive; for instance, fewer percentage points could be reported. However, these would be the domain of another study. The present study was designed to employ a specified pseudo-random number generator, a specified number of iterations, and a specified output format, and to test the reliability under these conditions. Furthermore, refinement of relianility is not considered by the present investigator to have the highest priority among problems to be considered in improving the simulation because the effects were relatively minor.

The effects of using the simulation rather than direct calculation were also relatively minor. (See Table 4.6.) For the calculations both for the first six years and for all twelve years, the arithmetic means are slightly smaller than those in the reliability test and the algebraic means are slightly larger. The fact that the means are so similar suggests that the differences between the simulated and nonsimulated predictions can be accounted for by the lack of reliability in the simulations. However, the fact that the algebraic means

are greater than those computed for the reliability may indicate a directional bias in the simulation, but the test is not definitive on this point. A bias might be attributed to lack of normality in the simulated distributions; in this case, the 0.50 figures would not be expected to conform exactly to the single figure predictions.

The number of times single figure multivariable predictions fell within the various cumulative probabilities of the simulated distributions were tabulated. Ten of the 144 multivariable prediction figures were the same as the 0.50 probability figures of the distributions produced by simulation using symmetrical data. One hundred twenty-nine of the single figures fell on or between the 0.40 and 0.60 figures, representing the points below which are 40 percent and 60 percent, respectively, of the figures in the distribution. All of the single figures were between the 0.30 and 0.70 probability figures.

Further studies could attempt to improve correspondence between simulated and nonsimulated results by improving reliability or by searching for a cause of the possible directional bias. Or they could concentrate on changing the output format so that the output does not give the impression

of accuracy greater than the simulation can provide. Or they could study what effect a lack of normality of the simulated distributions has on the correspondence of the single figure prediction to various points in the distributions. However, in the opinion of the present investigator, these studies do not have highest priority since the difference effects are relatively small.

Differences which are not as small are revealed by comparing output from the simulations using skewed and symmetrical data. The use of skewed data may make as much as 4.38 per cent difference. (See Table 4.6.) The number of times the single figure multivariable predictions fell within the various probability levels of the distributions using skewed data showed that single figures fell 67 times within the 0.20-0.29 figures and 55 times within the 0.30-0.39 figures. All of the single figure predictions fell within the 0.10 and 0.69 figures. The differences in these results and those produced by the symmetrical data reflect the larger enrollment figures produced by the skewed data. The impact of the skewed data seems to be large enough to suggest that other statistical models should be tried; the simulation developed in the present study employs beta distributions which are transformed into normal distributions.

However, it should be noted that the impact of skewed data might hot be so great with ordinary use of the simulation. The high and low estimates were set arbitrarily by the investigator in this study. In the set of skewed data, all of the variables were skewed in the same direction; use of authentic high and low estimates would probably produce a set of data in which some of the variables were not skewed and others were skewed in opposite directions.

Chapter V

Summary and Conclusions

The adequacy of enrollment predictions can have a substantial influence on the quality of educational programs. Difficulties associated with making adequate predictions are the extent of unpredictability of the phenomena, the inaccuracy of the prediction methods, and the lack of adequate communication between the forecaster and the user of the forecasts. Indeed, it is the uncertainties associated with the forecast figures that may be the most difficult to communicate. The present study is an attempt to develop a systematic means by which a forecaster can assess and express the uncertainties involved in the predictions. A method of preparing probability distributions of enrollment predictions was developed. The use of probability distributions has the advantage of presenting probabilistic information in a manner which is not as unwieldly as lists probabilities for various contingencies. Another advantage is that it integrates probabilistic information into the numerical and graphical presentation of predictions, making the message of unpredictability more forceful than in the words of warning often tucked into footnotes after pages of numbers and tables.

In the present study a basic method for single figure predictions was modified to accommodate probabilities. The multivariable method was chosen for this purpose since it allows the forecaster the freedom to deviate from simple projections and to make probability statements about individual variables. Monte Carlo computer simulation was chosen as the means for combining the probabilistic data to form the probability distributions.

Significance tests were performed to investigate predictive validity of the prediction method and reliability and concurrent validity of the simulation output. A comparison of the predictive validity of the multivariable method with that of the percentage of survival method in one school system showed greater accuracy for the multivariable method. However, other studies should be conducted to determine whether or not these results can be replicated and to compare the multivariable method with methods other than the percentage of survival method. The tests of reliability and concurrent validity showed evidence of undesired noise effects, but there is evidence to show that the effects of lack of reliability and lack of agreement between simulated and nonsimulated results are actually minimal. However, the lack of concurrent validity as measured by the effect of skewed input may be

large enough to justify reconsideration of the statistical distributions used to describe the variable estimates. This can best be determined by using data in which the high and low estimates are not arbitrarily set, as they were in the present study, but are actual estimates of probabilities by persons with sufficient information to make such estimates.

The major contribution of the study was not the information gained from testing the hypotheses, but the development of a method of predicting school enrollments in terms of probabilities. The development involved the choice of the multivariable method as the prediction model and the choice of computer simulation as the method of combining probabilistic estimates; these choices were made after reviewing the literature on enrollment prediction, population prediction, computer simulation, and other methods of handling estimates of probabilities. Other decisions that had to be made in the course of the development were the exact specifications of the multivariable model, the type of output to be produced, the type of input to be required, the form of the distributions for the probability estimates in the input, the development of the computer program, and the kinds of statistical tests

to be performed on the output. The outcome and rationale for these decisions are explained in Chapters III and IV. Other actions could have been taken at many of these decision points, but the present investigator felt that it would be more useful to prepare one complete model that could begin to be used than to conduct separate studies for each of the decisions that had to be made.

As a result of the developmental process, the computer programs and instructions are ready for use although it is suggested that they be employed with a heuristic approach, seeking more information about such factors as predictive validity. The development of the model and its use with trial data served to illustrate possible problems with the model. A question which was raised by the results of the simulation was that of independence of the variables.

The lack of independence among the variables is one explanation for the fact that approximately 40 percent of the actual Brockton enrollments fell outside the 0.05 and 0.95 probability levels. (See Table 5.1, p.137.) One should remember that the distributions produced by the simulations are dependent upon arbitrarily chosen high and low estimates; those estimates were, in general, chosen to be plus and minus three standard deviations from the mean for the

TABLE 5.I

NUMBER OF TIMES THE ACTUAL ENROLLMENTS EXCEEDED THE .05 AND .95 PROBABILITY POINTS OF THE PREDICTIONS

YEAR	RANDOM #1	RANDOM #2
1964	8	7
1965	4	4
1966	3	3
1967	6	7
1968	4	5
1969	3	3

previous few years. If the variables are truly independent, this indicates a large departure from past trends in Brockton. Alternatively, it could be hypothesized that the arbitrarily set high and low estimates were not inadequate, but that the assumption of independence constricted the simulation distributions. Even if the variables are adequately independent, there is probably dependence among the years and grades, i.e., migration rates in one grade or year probably correlate with migration rates in other grades or years, but the model requires an assumption of independence even across grades and years for the same variable.

Perhaps the most serious problems with the simulation model in its present form are this assumption of independence and the assumption of a beta distribution of estimates with the subsequent transformation into a normal distribution. Perhaps the most serious question about the attempt to use probability estimates in enrollment prediction is whether or not adequate probability estimates can be made. More research needs to be done on the basic problem of applying probabilities to enrollment predictions: both enrollment prediction methods and methods of estimating probabilities need to be validated and perfected. Since a primary purpose of the use of probability distributions of enrollment is to aid communication between the

forecaster and the user, studies of the effectiveness of various formats of presentation need to be conducted. Until such advances are made, the present study provides a simulation model for enrollment prediction whose use is not discouraged if the user keeps in mind the reservations stated by the investigator. Its use or the use of a modification of the model is encouraged if the user plans to take advantage of the opportunity to add to the research and validation data and to suggest improvements.

The model developed here is considered a prototype of models which could be developed. One might want to predict the enrollment figures for a state as a whole or for individual schools within a district. One might want predictions separately by race or scioeconomic class. An ungraded school might need predictions based on categories other than grade level, perhaps categories designating progress in relation to curriculum goals, so that demands on specialized teachers, equipment, and facilities might be anticipated. A unique contribution of the prototype for enrollment prediction is that it requires the user to examine the various parts of the system, to assess probabilities for these parts, and to use these probabilities to determine probabilistic information

about the operation of the system as a whole. It is the demonstration of this methodology in a workable model for computer simulation which gives this study its significance.

REFERENCES

American Association of School Administrators. Twenty-seventh Yearbook. American School Buildings. Washington, D.C.: American Association of School Administrators, 1947.

Armitage, Peter, and Smith, Cyril. The Development of Computable Models of the British Educational System and Their Possible Uses. Mathematical Models in Educational Planning, 159-206. Paris: Organization for Economic Co-operation and Development, 1967.

Bartos, Otomar J. Simple Models of Group Behavior. New York: Columbia University Press, 1967.

Arthur D. Little, Inc. (with Earl R. Flansburgh and Associates). A Development Program for Quality Education in Quincy. Cambridge, Mass.: Arthur D. Little, Inc.,February, 1966.

Beshers, James E. (Ed.) Computer Methods in the Analysis of Large-scale Social Systems. Proceedings of a conference held at the Joint Center for Urban Studies of the Massachusetts Institute of Technology and Harvard University in Cambridge, Massachusetts, October 19-21, 1964. Cambridge, Mass.: Joint Center for Urban Studies,1965.

Brackett, James W. Some Computer Applications to Population Projections and Demographic Analysis. In James E. Beshers (Ed.), Computer Methods in the Analysis of Large-scale Social Systems, 60-67. Cambridge, Mass.: Joint Center for Urban Studies, 1965.

Brockton School Department. Your Schools. Annual report of the School Department. Brockton, Mass.: Brockton School Department, 1959-1969.

Brown, Roscoe C.,Jr. Predicting School Enrollments. New York: Center for School Services, School of Education, New York University, 1961.

Burford, Roger L. Probability Projections of Rates of Net Migration for Southern Countries and Other Applications of Markov Chains. Louisiana Business Bulletin, 25(1) (July, 1966).

Cashwell, E. D. and Everett, C. J. A Practical Manual on the Monte Carlo Method for Random Walk Problems. New York: Pergamon Press, 1959.

Center for Field Studies, Harvard University Graduate School of Education. Brockton's Schools, 1964. Cambridge, Mass.: Center for Field Studies, 1964.

Center for Field Studies, Harvard University Graduate School of Education. The Children's Future School Buildings. Cambridge, Mass.: Center for Field Studies, 1953.

Center for Field Studies, Harvard University Graduate School of Education. Education for Pittsburgh. Cambridge, Mass.: Center for Field Studies, 1966.

Center for Field Studies, Harvard University Graduate School of Education. Salem School Buildings Needs. Cambridge, Mass.: Center for Field Studies, 1956.

Center for Field Studies, Harvard Univeristy Graduate School of Education. Schools for Hartford. Cambridge, Mass.: Center for Field Studies, n.d.

Center for Field Studies, Harvard University Graduate School of Education. A Valley and a Decision. Cambridge, Mass.: Center for Field Studies, 1954.

Citizens' Advisory Committee on School Needs, Abilene Independent School District. Unpublished report. Abilene, Texas: Citizens' Advisory Committee on School Needs, January 1, 1960.

Collins, George J. and Langston, La Moine. Guesstimating Future School Enrollments. American School Board Journal, 143 (December, 1961), 10-12.

Curtiss, J. H. Sampling Methods Applied to Differential and Difference Equations. Proceedings of Seminar on Scientific Computation, 87-109. Endicott, New York: International Business Machines Corporation, 1949.

Daniel, K. Fred. Computer Based System for Assisting in Planning. Memorandum to deans and directors of teacher education in Florida. Tallahassee: State of Florida Department of Education, October 1, 1969.

Demeny, Paul. Discussion. In James E. Beshers (Ed.), Computer Methods in the Analysis of Large-scale Social Systems, 68-70. Cambridge, Mass.: Joint Center for Urban Studies, 1965.

Derthick, Lawrence. Statement made before the Subcommittee on General Education, Committee on Education and Labor, U.S. House of Representatives, 85th Cong., 1st sess., February 6, 1957.

Forecasting Institute of the Swedish Central Bureau of Statistics, The. Projection Models of the Swedish Educational System. In Mathematical Models in Educational Planning, 95-124. Paris: Organization for Economic Co-operation and Development, 1967.

Gottlieb, Abe. A Planning Approach to School Enrollment Forecasts. American School Board Journal, 128(2) (February, 1954), 68-69, 110.

Greenawalt, George J. and Mitchell, Donald P. Predicting School Enrollments. Cambridge, Mass.: New England School Development Council, 1966.

Griffin, Mary and Schmitt, John. A Monte Carlo Model for the Prediction of Public School Enrollments. Paper presented at the 1966 Research Convocation of the Educational Research Association of New York State and the New York State Education Department, held at Albany, New York, November 14, 1966.

Griffith, William Jess. Variables Affecting Public School Enrollment Change. Unpublished Ph.D. thesis, The Ohio State University, 1967.

Hammersley, J. M., and Handscomb, D.C. Monte Carlo Methods. New York: John Wiley & Sons, Inc., 1964.

Hastings, Cecil. Approximations for Digital Computers. Princeton: Princeton University Press, 1955.

Hays, William L. Statistics, New York: Holt, Rinehart & Winston, 1963.

Hodge, Patricia Leavey, and Hauser, Philip M. The Challenge of America's Metropolitan Population Outlook--1960 to 1985. Prepared for the consideration of The National Commission on Urban Problems. Research Report No. 3. Washington, D.C.: U.S. Government Printing Office, 1968.

Hodgman, Charles D. (Ed.) Standard Mathematical Tables, 12th ed. Cleveland: Chemical Rubber Publishing Co., 1959.

Hunt, Herold, C., et al. Watertown: The Education of its Children. Cambridge, Mass.: Center for Field Studies, Harvard University Graduate School of Education, 1967.

International Business Machines Corporation. System/360 Scientific Subroutine Package, (360A-CM-03x) Version III, Application Description. 6th ed. New York: IBM Corporation, 1968.

International Business Machines Corporation. Data Processing Techniques: Random Number Generation and Testing. GC20-8011-0. 1st. ed. (Reprint December, 1969). New York: IBM Corporation, 1959.

Isard, Walter. et al. Methods of Regional Analysis: An Introduction to Regional Science. Cambridge, Mass.: The M.I.T. Press, 1960.

Jaffe, A. J. Progress Report on U.S. Office of Education Bureau of Research No. 7-1253, Handbook of Statistical Methods for Making School Enrollment Projections. April 29, 1968.

Johnson, Howard M., et al. Englewood Schools: A Study of the Englewood Public Schools, Englewood, New Jersey. Cambridge, Mass.: Center for Field Studies, Harvard University Graduate School of Education, 1965.

Johnson, Howard. Interlake High School, Bellevue, Washington. Letter, September 24, 1969.

Kendall, Maurice G. and Buckland, William R. A Dictionary of Statistical Terms. New York: Hafner Publishing Co., 1957.

Kendall, Maurice G. and Stuart, Alan. The Advanced Theory of Statistics. Vol. 1. Distribution Theory. (2nd ed.), New York: Hafner Publishing Co., 1963.

Lambourne, Simon. Network Analysis in Project Management. London: Industrial and Commerical Techniques, Ltd., 1967.

Larson, Knute G. and Strevell, Wallace H. How Reliable Are Enrollment Forecasts? The School Executive, 71 (February, 1952), 65-68.

MacConnell, James D. Planning for School Buildings. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1957.

McCracken, Daniel D. The Monte Carlo Method. Scientific American, 193(5) (May, 1955), 90-95.

Malcolm, Donald G. System Simulation--a Fundamental Tool for Industrial Engineering. The Journal of Industrial Engineering, 9 (May-June, 1958), 177-187.

Marshall, A. W. An Introductory Note. In H. A. Meyers (Ed.), Symposium on Monte Carlo Methods. New York: John Wiley & Sons, Inc., 1956.

Marshall, John E. School Building Needs: Methuen, Mass. Unpublished report, October 1968.

Martin, Francis F. Computer Modeling and Simulation. New York: John Wiley & Sons, Inc., 1968.

Massachusetts Department of Commerce and Development. City and Town Monograph, City of Brockton. Boston: Massachusetts Department of Commerce and Development, November, 1967.

Metcalf and Eddy, Inc., Engineers and Planners. Long-range Comprehensive Town and Financial Plan for Lexington, Massachusetts, Part II--Long-range Physical Facilities Plans. Boston: Metcalf and Eddy, Inc., 1968.

Meyer, Herbert A. (Ed.) Symposium on Monte Carlo Methods. From a symposium sponsored by Wright Air Development Center of the Air Research and Development Command, held at the University of Florida, March 16-17, 1954. New York: John Wiley & Sons, Inc., 1954.

The Nebraska Co-ordinating Council for Public Higher Education, et al. A Demographic Study for Nebraska Higher Education. Part I. Projections: Elementary - Secondary Enrollments through 1975. Lincoln: The Nebraska Co-ordinating Council for Public Higher Education, June, 1967.

Orcutt, Guy H. Data Needs for Computer Simulation of Large-scale Social Systems. In James E. Beshers (Ed.), Computer Methods in the Analysis of Large-scale Social Systems, 189-198. Cambridge, Mass.: Joint Center for Urban Studies, 1964.

Orcutt, Guy H., et al. Microanalysis of Socioeconomic Systems: A Simulation Study. New York: Harper & Bros., 1961.

Orcutt, Guy H. Simulation of Economic Systems. In Harold Guetzkow (Ed.), Simulation in Social Science: Readings. Englewood Cliffs, N.J.: Prentice Hall, 1962.

Perrin, E. B., and Sheps, Mindel C. Human Reproduction: a Stochastic Process. Biometrics, 20 (March, 1964), 28-45.

Perrin, E. B., and Sheps, Mindel C. A Monte Carlo Investigation of a Human Fertility Model. Paper presented before the Statistics Section, American Public Health Association, November, 1963.

PERT Orientation and Training Center. PERT Fundamentals Vol. III -- Workbook. Washington, D.C.: U.S. Government Printing Office, 1963.

Peters, Stefan. Working Memorandum SP-2 Case No.: 1-8324. Unpublished memorandum. Cambridge, Mass.: Arthur D. Little, Inc., 1969.

Population and Housing Committee, Citizens' Committee for the Lexington Long-range Study. Unpublished report. Lexington, Mass.: Population and Housing Committee, November 1, 1966.

Rosenberg, Harry M. Estimating Changes in Population. Battelle Technical Review, April, 1967), 2-6.

Sargent, Cyril G., et al. Boston Schools, 1962. Boston: Boston Redevelopment Authority, May, 1962.

Scammon, Richard. Projecting Student Enrollment. Overview, 3 (December, 1962), 39-41.

Sheps, Mindel C. Application of Probability Models to the Study of Patterns of Human Reproduction. In Mindel C. Sheps and Jeanne Clare Ridley (Eds.), Public Health and Population Change: Current Research Issues, 307-332. Pittsburgh: University of Pittsburgh Press, 1965.

Sheps, Mindel C. and Perrin, E. B. Changes in Birth Rate as a Function of Contraceptive Effectiveness: Some Applications of a Stochastic Model. American Journal of Public Health, 53 (July, 1963), 1031-1046.

Siegel, Jacob S. Forecasting the Population of Small Areas. Land Economics, 29(1) (February, 1953).

Siegel, Sidney. Nonparametric Statistics for the Behavioral Sciences. New York: McGraw-Hill, Inc., 1956.

Simon, Kenneth A. and Fullam, Marie G. Projections of Educational Statistics to 1976-77. Washington, D.C.: U.S. Government Printing Office, 1968.

Stanbery, Van Beuren. Better Population Forecasting for Areas and Communities. U.S. Department of Commerce, Domestic Commerce Series, No. 32, Washington, D.C.: U.S. Government Printing Office, 1952.

Strevell, Wallace H. Techniques of Estimating Future Enrollment. American School Board Journal, (1952), 35-38.

Thompson, Warren S., and Whelpton, P.K. Population Trends in the United States. New York: McGraw-Hill, Inc., 1933.

Thonstad, Tore. A Mathematical Model of the Norwegian Educational System. Mathematical Models in Educational Planning, 125-158. Paris: Organization for Economic Co-operation and Development, 1967.

UNESCO. Computer to Aid Long-range Planning of School Systems. School and Society, 194 (March 5, 1966), 122, 129.

U.S. Department of the Navy, Special Projects Office, Bureau of Naval Weapons. PERT Summary Report, Phase I. Washington, D.C.: U.S. Government Printing Office, 1958.

Whelpton, Pascal K. Cohort Fertility, Native White Women in the United States. Princeton: Princeton University Press, 1954.

Whitla, Dean K. On the Prediction of School Enrollment. Unpublished paper. Cambridge, Mass.: Graduate School of Education, Harvard University, December, 1954.

Zabrowski, Edward K. The Dynamod Model of Student and Teacher Population Growth. Socioeconomic Planning Sciences, 2 (April, 1969), 455-464.

APPENDIXES

APPENDIX A

Program MAIN

```
0001        DIMENSION TPROJ(100), BPROJ(100), GPROJ(100), ENROLL(2,13),
           2BIRTHS(2,5), DEATHS(2,5), ABIRTH(6,11), BBIRTH(6,11),CBIRTH(6,11),
           3DBIRTH(6,11), XBIRTH(6,11), SBIRTH(6,11), AGIRLS(6,11),
           4DGIRLS(6,11), CGIRLS(6,11), XGIRLS(6,11), SGIRLS(6,11),
           5ADEATH(11), BDEATH(11), CDEATH(11), DDEATH(11), XDEATH(11),
           6SDEATH(11), APREMI(15), BPREMI(15), CPREMI(15), DPREMI(15),
           7XPREMI(15), SPREMI(15), AMIGRA(13,15), BMIGRA(13,15),
           8CMIGRA(13,15), DMIGRA(13,15), XMIGRA(13,15), SMIGRA(13,15),
           9APRIVS(15), BPRIVS(15), CPRIVS(15), DPRIVS(15), XPRIVS(15),
           ASPRIVS(15), ATRANS(13,15), BTRANS(13,15), CTRANS(13,15),
           BDTRANS(13,15), XTRANS(13,15), STRANS(13,15), AHOLDS(13,15),
           CBHOLDS(13,15), CHOLDS(13,15), DHOLDS(13,15), XHOLDS(13,15),
           DSHOLDS(13,15), AINSTI(13,15), BINSTI(13,15), CINSTI(13,15),
           EDINSTI(13,15), XINSTI(13,15), SINSTI(13,15), ADROPS(5,15),
           FBDROPS(5,15), CDROPS(5,15), DDROPS(5,15), XDROPS(5,15),
           GSDROPS(5,15), TITLE(20), FORM1(20), FORM2(20), FORM3(20),
           HFORM4(20), FORM5(20), FORM6(20), FORM7(20), FORM8(20), FORM9(20),
           IFORM10(20), FORM11(20), FORM12(20), FORM13(20), BIRTH(6),
           JRETAIN(13,15), RHOLDS(100)
0002        DIMENSION TPREVY(13,100), BPREVY(13,100), TPRESY(13,100),
           2BPRESY(13,100), TOTBIR(6), BOYBIR(6)
0003        READ (5,1) TITLE
0004        READ(5,2) IYEAR, IGRADE, ISEX, IDATE, IX
0005        READ(5,3) FORM1, FORM2, FORM3, FORM4, FORM5, FORM6, FORM7, FORM8,
           2FORM9, FORM10, FORM11, FORM12, FORM13
0006      1 FORMAT (20A4)
0007      2 FORMAT (8X, I2, 3X, I2, 4X, I1, 1X, I4, 1X, I9)
0008      3 FORMAT (20A4)
0009        IF (ISEX .EQ. 1) GO TO 444
0010        READ (5,FORM1) (ENROLL(1,J), J = 1,IGRADE)
0011        IF (IGRADE .EQ. 12) LIMIT = 5
0012        IF (IGRADE .EQ. 13) LIMIT = 4
0013        READ (5,FORM2) (BIRTHS(1,J), J = 1,LIMIT)
0014        READ (5,FORM3) (DEATHS(1,J), J = 1,LIMIT)
0015        ISTOP = IYEAR - LIMIT
0016        READ (5,FORM4) ((ABIRTH(I,J), BBIRTH(I,J), CBIRTH(I,J), I = 1,6),
           2J = 1,ISTOP)
0017        READ (5,FORM5) ((AGIRLS(I,J), BGIRLS(I,J), CGIRLS(I,J), I = 1,6),
           2J=1,ISTOP)
0018        READ (5,FORM6) (ADEATH(J), BDEATH(J), CDEATH(J), J = 1,ISTOP)
0019        READ (5,FORM7) (APREMI(J), BPREMI(J), CPREMI(J), J = 1,IYEAR)
0020        READ (5,FORM8) ((AMIGRA(I,J), BMIGRA(I,J), CMIGRA(I,J),
           2I=2,IGRADE), J = 1,IYEAR)
0021        READ (5,FORM9) (APRIVS(J), BPRIVS(J), CPRIVS(J), J = 1,IYEAR)
0022        READ (5,FORM10) ((ATRANS(I,J), BTRANS(I,J), CTRANS(I,J),
           2I = 2,IGRADE), J = 1,IYEAR)
0023        READ (5,FORM11) ((AHOLDS(I,J), BHOLDS(I,J), CHOLDS(I,J),
           2I=1,IGRADE), J = 1,IYEAR)
```

FILMED FROM BEST AVAILABLE COPY

Program MAIN (continued)

```
0024          READ (5,FORM12) ((AINSTI(I,J), BINSTI(I,J), CINSTI(I,J),
             2I=2,IGRADE), J=I,IYEAR)
0025          READ (5,FORM13) ((ADROPS(I,J), BDROPS(I,J), CDROPS(I,J), I =1,5),
             2J = 1,IYEAR)
0026          GO TO 222
0027      444 READ(5,FORM1) ((ENROLL(I,J), I= 1,2), J = 1,IGRADE)
0028          IF (IGRADE .EQ. 12) LIMIT = 5
0029          IF (IGRADE .EQ. 13) LIMIT = 4
0030          READ(5,FORM2) ((BIRTHS(I,J), I = 1,2), J = 1,LIMIT)
0031          READ(5,FORM3) ((DEATHS(I,J), I = 1,2), J = 1,LIMIT)
0032          ISTOP = IYEAR - LIMIT
0033          READ (5,FORM4) ((ABIRTH(I,J), BBIRTH(I,J), CBIRTH(I,J),
             2DBIRTH(I,J), I = 1,6), J = 1,ISTOP)
0034          READ (5,FORM5) ((AGIRLS(I,J), BGIRLS(I,J), CGIRLS(I,J), I = 1,6),
             2J = 1,ISTOP)
0035          READ (5,FORM6) (ADEATH(J), BDEATH(J), CDEATH(J), DDEATH(J),
             2J = 1,ISTOP)
0036          READ (5,FORM7) (APREMI(J), BPREMI(J), CPREMI(J), DPREMI(J),
             2J=1,IYEAR)
0037          READ (5,FORM8) ((AMIGRA(I,J), BMIGRA(I,J), CMIGRA(I,J),
             2DMIGRA(I,J), I = 2,IGRADE), J = 1,IYEAR)
0038          READ (5,FORM9) (APRIVS(J), BPRIVS(J), CPRIVS(J), DPRIVS(J),
             2J=1,IYEAR)
0039          READ (5,FORM10) ((ATRANS(I,J), BTRANS(I,J), CTRANS(I,J),
             2DTRANS(I,J), I =2,IGRADE), J = 1,IYEAR)
0040          READ (5,FORM11) ((AHOLDS(I,J), BHOLDS(I,J), CHOLDS(I,J),
             2DHOLDS(I,J), I = 1,IGRADE), J = 1,IYEAR)
0041          READ (5,FORM12) ((AINSTI(I,J), BINSTI(I,J), CINSTI(I,J),
             2DINSTI(I,J), I = 2,IGRADE), J = 1,IYEAR)
0042          READ (5,FORM13) ((ADROPS(I,J), BDROPS(I,J), CDROPS(I,J),
             2DDROPS(I,J), I=1,5), J=1,IYEAR)
C             THIS PART OF THE PROGRAM PRINTS INPUT AS A CHECK, COMPUTES AND
C             PRINTS MEANS AND STANDARD DEVIATIONS.
0043      222 WRITE(6,4) TITLE, IYEAR, IGRADE, ISEX, IDATE, IX, FORM1, FORM2,
             2FORM3, FORM4, FORM5, FORM6, FORM7, FORM8, FORM9, FORM10, FORM11,
             3FORM12, FORM13
0044        4 FORMAT (1H1,20A4//1X,'NUMBER OF YEARS',5X,I2/1X,
             2'NUMBER OF GRADES',5X,I2/1X,'SEX OPTION',5X,I1/1X,
             3'BEGINNING YEAR',5X,I4/1X,'SEED',5X,I9//1X,'INPUT FORMATS'//
             4(1X,20A4))
0045          IF (ISEX .EQ. 1) GO TO 13131
0046          WRITE (6,5) (J,ENROLL(1,J), J = 1,IGRADE)
0047        5 FORMAT (1H1,'ENROLLMENT BY GRADE IN BASE YEAR'//(1X,I2,5X,F10.0))
0048          WRITE (6,6) (J,BIRTHS(1,J), J = 1,LIMIT)
0049        6 FORMAT (1H1,'HISTORICAL BIRTHS BY YEAR OF SIMULATION'//
             2(1X,I2,5X,F10.0))
0050          WRITE (6,7) (J,DEATHS(1,J), J = 1,LIMIT)
0051        7 FORMAT (1H1,'PRESCHOOL DEATHS BY YEAR OF SIMULATION'//
```

Program MAIN (continued)

```
                2(1X,I2,5X,F10.0))
0052             GO TO 1313
0053        13131 WRITE(6,505) (J, ENROLL(1,J), ENROLL(2,J), J = 1,IGRADE)
0054          505 FORMAT (1H1,'ENROLLMENT BY GRADE IN BASE YEAR'//13X,'TOTAL',10X,
                2'BOYS'//(1X,I2,2(5X,F10.0)))
0055             WRITE(6,606) (J, BIRTHS(1,J), BIRTHS(2,J), J = 1,LIMIT)
0056          606 FORMAT (1H1,'HISTORICAL BIRTHS BY YEAR OF SIMULATION'//9X,'TOTAL',
                25X,'BOYS'//(1X,I2,2(5X,F10.0)))
0057             WRITE (6,707) (J, DEATHS(1,J), DEATHS(2,J), J = 1,LIMIT)
0058          707 FORMAT (1H1,'PRESCHOOL DEATHS BY YEAR OF SIMULATION'//
                214X,'TOTAL',10X,'BOYS'//(1X,I2,2(5X,F10.0)))
0059         1313 WRITE (6,8)
0060            8 FORMAT (1H0,'PREDICTIONS FOR BIRTH RATE BY AGE GROUP'/)
0061             IF (ISEX .EQ. 1) WRITE (6,909)
0062             IF (ISEX .EQ. 0) WRITE (6,9)
0063          909 FORMAT (1H ,'YEAR',2X,'LEVEL',2X,'HIGH ESTIMATE',3X,'LIKELY ESTIMA
                2TE',3X,'LOW ESTIMATE',5X,'MEAN',5X,'STANDARD DEVIATION',5X,'PROPOR
                3TION BOYS'/)
0064            9 FORMAT (1H ,'YEAR',2X,'LEVEL',2X,'HIGH ESTIMATE',3X,
                2'LIKELY ESTIMATE',3X,'LOW ESTIMATE',5X,'MEAN',5X,
                3'STANDARD DEVIATION'/)
0065             DO 400 J = 1,ISTOP
0066             DO 400 I = 1,6
0067             XBIRTH(I,J) = (ABIRTH(I,J) + 4. * BBIRTH(I,J) + CBIRTH(I,J))/6.0
0068             SBIRTH(I,J) = (ABIRTH(I,J) - CBIRTH(I,J))/6.0
0069             IF (ISEX .EQ. 1) WRITE(6,404) J, I, ABIRTH(I,J), BBIRTH(I,J),
                2CBIRTH(I,J), XBIRTH(I,J), SBIRTH(I,J), DBIRTH(I,J)
0070             IF (ISEX .EQ. 0) WRITE(6,40) J, I, ABIRTH(I,J), BBIRTH(I,J),
                2CBIRTH(I,J), XBIRTH(I,J), SBIRTH(I,J)
0071          404 FORMAT (1H ,I2,5X,I2,3X,3(F10.1,5X),F10.2,5X,F10.2,15X,F6.3)
0072           40 FORMAT (1H ,I2,5X, I2, 3X, 3(F10.1,5X), F10.2, 5X, F10.2)
0073          400 CONTINUE
0074             WRITE (6,51)
0075           51 FORMAT (1H0,'PREDICTIONS FOR FEMALES BY AGE GROUP'/)
0076             WRITE (6,9)
0077             DO 500 J = 1,ISTOP
0078             DO 500 I = 1,6
0079             XGIRLS(I,J) = (AGIRLS(I,J) + 4. * BGIRLS(I,J) + CGIRLS(I,J))/6.0
0080             SGIRLS(I,J) = (AGIRLS(I,J) - CGIRLS(I,J))/6.0
0081             WRITE (6,40) J,I,AGIRLS(I,J), BGIRLS(I,J), CGIRLS(I,J),
                2XGIRLS(I,J), SGIRLS(I,J)
0082          500 CONTINUE
0083             WRITE (6,61)
0084           61 FORMAT (1H0,'PREDICTIONS OF PERCENTAGES OF PRESCHOOL DEATHS'/)
0085             IF (ISEX .EQ. 1) WRITE(6,1010)
0086             IF (ISEX .EQ. 0) WRITE(6,10)
0087         1010 FORMAT (1H ,'YEAR',2X,'HIGH ESTIMATE',3X,'LIKELY ESTIMATE',3X,
                2'LOW ESTIMATE',5X,'MEAN',5X,'STANDARD DEVIATION',5X,
```

Program MAIN (continued)

```
                3'PROPORTION BOYS'/)
0088          10 FORMAT (1H ,'YEAR',2X,'HIGH ESTIMATE',3X,'LIKELY ESTIMATE',3X,
                2'LOW ESTIMATE',5X,'MEAN',5X,'STANDARD DEVIATION'/)
0089              DO 600 J = 1,ISTOP
0090              XDEATH(J) = (ADEATH(J) + 4.*BDEATH(J) + CDEATH(J))/6.0
0091              SDEATH(J) = (ADEATH(J) - CDEATH(J))/6.0
0092              IF (ISEX .EQ. 1) WRITE(6,6060) J, ADEATH(J), BDEATH(J), CDEATH(J),
                2XDEATH(J), SDEATH(J), DDEATH(J)
0093              IF (ISEX .EQ. 0) WRITE(6,60) J, ADEATH(J), BDEATH(J), CDEATH(J),
                2XDEATH(J), SDEATH(J)
0094        6060 FORMAT (1H ,I2,5(5X,F10.7),12X,F6.3)
0095          60 FORMAT (1H ,I2,5(5X,F10.7))
0096         600 CONTINUE
0097              WRITE (6,71)
0098          71 FORMAT (1H0,'PREDICTIONS OF PRESCHOOL MIGRATION'/)
0099              IF (ISEX .EQ. 1) WRITE(6,1010)
0100              IF (ISEX .EQ. 0) WRITE(6,10)
0101              DO 700 J = 1, IYEAR
0102              XPREMI(J) = (APREMI(J) + 4. * BPREMI(J) + CPREMI(J))/6.0
0103              SPREMI(J) = (APREMI(J) - CPREMI(J))/6.0
0104              IF (ISEX .EQ. 1) WRITE(6,7070) J, APREMI(J), BPREMI(J),
                2CPREMI(J), XPREMI(J), SPREMI(J), DPREMI(J)
0105              IF (ISEX .EQ. 0) WRITE(6,70) J, APREMI(J), BPREMI(J), CPREMI(J),
                2XPREMI(J), SPREMI(J)
0106        7070 FORMAT (1H ,I2,3(5X,F10.1),2(5X,F10.2),12X,F6.3)
0107          70 FORMAT (1H ,I2,3(5X,F10.1),2(5X,F10.2))
0108         700 CONTINUE
0109              WRITE (6,81)
0110          81 FORMAT (1H0,'PREDICTIONS OF NET MIGRATION BY GRADE'/)
0111              IF (ISEX .EQ. 1) WRITE(6,909)
0112              IF (ISEX .EQ. 0) WRITE (6,9)
0113              DO 800 J = 1,IYEAR
0114              DO 800 I = 2,IGRADE
0115              XMIGRA (I,J) = (AMIGRA(I,J) + 4. * BMIGRA(I,J) + CMIGRA(I,J))/6.0
0116              SMIGRA(I,J) = (AMIGRA(I,J) - CMIGRA(I,J))/6.0
0117              IF (ISEX .EQ. 1) WRITE(6,404) J, I, AMIGRA(I,J), BMIGRA(I,J),
                2CMIGRA(I,J), XMIGRA(I,J), SMIGRA(I,J), DMIGRA(I,J)
0118              IF (ISEX .EQ. 0) WRITE(6,40) J, I, AMIGRA(I,J), BMIGRA(I,J),
                2CMIGRA(I,J), XMIGRA(I,J), SMIGRA(I,J)
0119         800 CONTINUE
0120              WRITE (6,91)
0121          91 FORMAT (1H0,'PREDICTIONS OF POTENTIAL LEVEL 1 STUDENTS ENROLLING
                2IN NON-PUBLIC SCHOOLS'/)
0122              IF (ISEX .EQ. 1) WRITE(6,1010)
0123              IF (ISEX .EQ. 0) WRITE (6,10)
0124              DO 900 J = 1,IYEAR
0125              XPRIVS(J) = (APRIVS(J) + 4. * BPRIVS(J) + CPRIVS(J))/6.0
0126              SPRIVS(J) = (APRIVS(J) - CPRIVS(J))/6.0
```

Program MAIN (continued)

```
0127          IF (ISEX .EQ. 1) WRITE (6,7070) J, APRIVS(J), BPRIVS(J),
             2CPRIVS(J), XPRIVS(J), SPRIVS(J), DPRIVS(J)
0128          IF (ISEX .EQ. 0) WRITE (6,70) J, APRIVS(J), 3PRIVS(J), CPRIVS(J),
             2XPRIVS(J), SPRIVS(J)
0129      900 CONTINUE
0130          WRITE (6,101)
0131      101 FORMAT (1H0,'PREDICTIONS OF TRANSFERS TO/FROM NON-PUBLIC SCHOOLS
             2BY GRADE'/)
0132          IF (ISEX .EQ. 1) WRITE (6,909)
0133          IF (ISEX .EQ. 0) WRITE (6,9)
0134          DO 1000 J = 1,IYEAR
0135          DO 1000 I = 2,IGRADE
0136          XTRANS(I,J) = (ATRANS(I,J) + 4. * BTRANS(I,J) + CTRANS(I,J))/6.0
0137          STRANS(I,J) = (ATRANS(I,J) - CTRANS(I,J))/6.0
0138          IF (ISEX .EQ. 1) WRITE(6,404) J, I, ATRANS(I,J), BTRANS(I,J),
             2CTRANS(I,J), XTRANS(I,J), STRANS(I,J), DTRANS(I,J)
0139          IF (ISEX .EQ. 0) WRITE (6,40) J, I, ATRANS(I,J), BTRANS(I,J),
             2CTRANS(I,J), XTRANS(I,J), STRANS(I,J)
0140     1000 CONTINUE
0141          WRITE (6,111)
0142      111 FORMAT (1H0,'PREDICTIONS OF PERCENTAGES RETAINED IN EACH GRADE'/)
0143          IF (ISEX .EQ. 1) WRITE (6,909)
0144          IF (ISEX .EQ. 0) WRITE (6,9)
0145          DO 1100 J = 1,IYEAR
0146          DO 1100 I = 1,IGRADE
0147          XHOLDS(I,J) = (AHOLDS(I,J) + 4. * BHOLDS(I,J) + CHOLDS(I,J))/6.0
0148          SHOLDS(I,J) = (AHOLDS(I,J) - CHOLDS(I,J))/6.0
0149          IF (ISEX .EQ. 1) WRITE (6,5656) J, I, AHOLDS(I,J),
             2BHOLDS(I,J), CHOLDS(I,J), XHOLDS(I,J), SHOLDS(I,J), DHOLDS(I,J)
0150          IF (ISEX .EQ. 0) WRITE (6,110) J, I, AHOLDS(I,J),
             2BHOLDS(I,J), CHOLDS(I,J), XHOLDS(I,J), SHOLDS(I,J)
0151     5656 FORMAT (1H ,I2,5X,I2,5(5X,F10.7),12X,F6.3)
0152      110 FORMAT (1H , I2, 5X, I2, 5(5X, F10.7))
0153     1100 CONTINUE
0154          WRITE (6,121)
0155      121 FORMAT (1H0,  'PREDICTIONS OF PERCENTAGE LOSS BY GRADE BECAUSE OF
             2DEATH OR INSTITUTIONALIZATION'/)
0156          IF (ISEX .EQ. 1) WRITE (6,909)
0157          IF (ISEX .EQ. 0) WRITE (6,9)
0158          DO 1200 J = 1,IYEAR
0159          DO 1200 I = 2, IGRADE
0160          XINSTI(I,J) = (AINSTI(I,J) + 4. * BINSTI(I,J) + CINSTI(I,J))/6.0
0161          SINSTI(I,J) = (AINSTI(I,J) - BINSTI(I,J))/6.0
0162          IF (ISEX .EQ. 1) WRITE (6,5656) J, I, AINSTI(I,J), BINSTI(I,J),
             2CINSTI(I,J), XINSTI(I,J), SINSTI(I,J), DINSTI(I,J)
0163          IF (ISEX .EQ. 0) WRITE (6,110) J, I, AINSTI(I,J), BINSTI(I,J),
             2CINSTI(I,J), XINSTI(2,J), SINSTI(I,J)
0164     1200 CONTINUE
```

Program Main (continued)

```
0165          WRITE (6,131)
0166      131 FORMAT (1HC,'PREDICTIONS OF PERCENTAGE OF DROPOUTS, GRADES 7 - 11'
             2/)
0167          IF (ISEX .EQ. 1) WRITE (6,909)
0168          IF (ISEX .EQ. 0) WRITE(6,9)
0169          DO 1300 J = 1,IYEAR
0170          DO 1300 I = 1,5
0171          XDROPS(I,J) = (ADROPS(I,J) + 4. * BDROPS(I,J) + CDROPS(I,J))/6.0
0172          SDROPS(I,J) = (ADROPS(I,J) - CDROPS(I,J))/6.0
0173          IF (ISEX .EQ. 1) WRITE (6,5656) J, I, ADROPS(I,J),
             2BDROPS(I,J), CDROPS(I,J), XDROPS(I,J), SDROPS(I,J),
             3DDROPS(I,J)
0174          IF (ISEX .EQ. 0) WRITE (6,110) J, I, ADROPS(I,J), BDROPS(I,J),
             2CDROPS(I,J), XDROPS(I,J), SDROPS(I,J)
0175     1300 CONTINUE
        C     THIS PART OF PROGRAM BEGINS ITERATIONS AND PROJECTS FIRST GRADE
        C     (OR KINDERGARTEN) ENROLLMENT FOR THE FIRST 5(OR 4) YEARS OF THE
        C     SIMULATION
0176          AM = 0.0
0177          S = 1.0
0178          DO 1234 I = 1,1000
0179     1234 CALL GAUSS (IX, S, AM, V)
0180          DO 11111 J = 1,IYEAR
0181          DO 22222 I = 1,IGRADE
0182          IF (I .GT. 1) GO TO 33333
0183          IF (J .GT. LIMIT) GO TO 44444
0184          DO 10001 M = 1,100
0185          TPROJ(M) = BIRTHS(1,J) - DEATHS(1,J)
0186          CALL GAUSS(IX,SPREMI(J),XPREMI(J),V1)
0187          TPROJ(M) = TPROJ(M) + V1
0188          CALL GAUSS (IX,SPRIVS(J),XPRIVS(J),V2)
0189          TPROJ(M) = TPROJ(M) - V2
0190          CALL GAUSS(IX,SHOLDS(I,J), XHOLDS(I,J),V3)
0191          RHOLDS(M) = V3
0192          IF (J .EQ. 1) TPROJ(M) = TPROJ(M) + V3 * ENROLL(1,1)
0193          IF (J .GT. 1) TPROJ(M) = TPROJ(M) + V3 * TPREVY(1,M)
0194          TPRESY(1,M) = TPROJ(M)
0195          IF (ISEX .EQ. 0) GO TO 10001
0196          BPROJ(M) = BIRTHS(2,J) - DEATHS(2,J)
0197          BPROJ(M) = BPROJ(M) + V1 * DPREMI(J)
0198          BPROJ(M) = BPROJ(M) - V2 * DPRIVS(J)
0199          IF (J .EQ. 1) BPROJ(M) = BPROJ(M) + V3 * DHOLDS(I,J) * ENROLL(2,1)
0200          IF (J .GT. 1) BPROJ(M) = BPROJ(M) + V3 * DHOLDS(1,J) *
             2BPREVY(1,M)
0201          BPRESY(1,M) = BPROJ(M)
0202    10001 CONTINUE
0203          IF (ISEX .EQ. 0) GO TO 98777
0204          DO 77777 M = 1,100
```

Program MAIN (continued)

```
      GPROJ(M) = TPROJ(M) - BPROJ(M)
77777 CONTINUE
93777 ICOUNT = 1
      CALL OUTPUT(I,J,TPROJ,ICOUNT,IGRADE,IDATE)
      IF (ISEX .EQ. 0) GO TO 22222
      ICOUNT = 2
      CALL OUTPUT (I,J,BPROJ,ICOUNT,IGRADE,IDATE)
      ICOUNT = 3
      CALL OUTPUT (I,J,GPROJ,ICOUNT,IGRADE,IDATE)
      GO TO 22222
C     THIS PART OF PROGRAM PROJECTS FIRST GRADE (OR KINDERGARTEN)
C     ENROLLMENT FOR THE REMAINING YEARS OF THE SIMULATION
44444 DO 20002 M = 1,100
      N = J - LIMIT
      TOTAL1 = 0.0
      TOTAL2 = 0.0
      DO 19 L = 1,6
      CALL GAUSS(IX, SBIRTH(L,N), XBIRTH(L,N),V)
      VV = V
      CALL GAUSS (IX,SGIRLS(L,N),XGIRLS(L,N),V)
      TOTBIR(L) = VV * V / 1000.00
      TOTAL1 = TOTAL1 + TOTBIR(L)
      IF (ISEX .EQ. 0) GO TO 19
      BOYBIR(L) = VV * DBIRTH(L,N) * V / 1000.00
      TOTAL2 = TOTAL2 + BOYBIR(L)
   19 CONTINUE
      CALL GAUSS(IX, SDEATH(N), XDEATH(N), V1)
      TPROJ(M) = TOTAL1 - V1 * TOTAL1
      CALL GAUSS(IX, SPREMI(J), XPREMI(J), V2)
      TPROJ(M) = TPROJ(M) + V2
      CALL GAUSS(IX, SPRIVS(J), XPRIVS(J),V3)
      TPROJ(M) = TPROJ(M) - V3
      CALL GAUSS(IX,SHOLDS(1,J), XHOLDS(1,J), V4)
      RHOLDS(M) = V4
      TPROJ(M) = TPROJ(M) + V4 * TPREVY(1,M)
      TPRESY(1,M) = TPROJ(M)
      IF (ISEX .EQ. 0) GO TO 20002
      BPROJ(M) = TOTAL2 - V1 * TOTAL2
      BPROJ(M) = BPROJ(M) + V2 * DPREMI(J)
      BPROJ(M) = BPROJ(M) - V3 * DPRIVS(J)
      BPROJ(M) = BPROJ(M) + V4 * DHOLDS(1,J) * BPREVY(1,M)
      BPRESY(1,M) = BPROJ(M)
20002 CONTINUE
      IF (ISEX. EQ. 0) GO TO 98764
      DO 70007 M = 1,100
      GPROJ(M) = TPROJ(M) - BPROJ(M)
70007 CONTINUE
98764 ICOUNT = 1
```

Program MAIN (continued)

```
      CALL OUTPUT (I,J,TPROJ,ICOUNT,IGRADE,IDATE)
      IF (ISEX .EQ. 0) GO TO 22222
      ICOUNT = 2
      CALL OUTPUT (I,J,BPROJ,ICOUNT,IGRADE,IDATE)
      ICOUNT = 3
      CALL OUTPUT(I,J,GPROJ,ICOUNT,IGRADE,IDATE)
      GO TO 22222
C     THIS PART OF PROGRAM PROJECTS ENROLLMENT FOR SECOND (OR FIRST)
C     GRADE THROUGH GRADE 12 (OR 13) FOR THE FIRST YEAR OF SIMULATION
33333 IF (J .GT. 1) GO TO 88888
      DO 30003 M = 1,100
      TPROJ(M) = ENROLL(1,I-1)
      CALL GAUSS(IX, SMIGRA(I,1), XMIGRA(I,1),V1)
      TPROJ(M) = TPROJ(M) + V1
      CALL GAUSS (IX,STRANS(I,1),XTRANS(I,1),V2)
      TPROJ(M) = TPROJ(M) + V2
      TPROJ(M) = TPROJ(M) - RHOLDS(M) * ENROLL(1,I-1)
      CALL GAUSS(IX, SHOLDS(I,J), XHOLDS(I,J), V3)
      RHOLDS(M) = V3
      TPROJ(M) = TPROJ(M) + V3 * ENROLL(1,I)
      CALL GAUSS(IX,SINSTI(I,1),XINSTI(I,1),V4)
      TPROJ(M) = TPROJ(M) - V4 * ENROLL(1,I-1)
      IF (IGRADE .EQ. 13 .AND. I .GT. 8) GO TO 33
      IF (IGRADE .EQ. 12 .AND. I .GT. 7) GO TO 33
      GO TO 6061
   33 L = I - IGRADE + 5
      CALL GAUSS(IX,SDROPS(L,1),XDROPS(L,1),V5)
      TPROJ(M) = TPROJ(M) - V5 * ENROLL(1,I-1)
 6061 TPRESY(I,M) = TPROJ(M)
      IF (ISEX .EQ. 0) GO TO 30003
      BPROJ(M) = ENROLL(2,I-1)
      BPROJ(M) = BPROJ(M) + V1  * DMIGRA(I,1)
      BPROJ(M) = BPROJ(M) + V2 * DTRANS(I,1)
      BPROJ(M) = BPROJ(M) - RHOLDS(M) * DHOLDS(I-1,J) * ENROLL(2,I-1)
      BPROJ(M) = BPROJ(M) + V3 * DHOLDS(I,J) * ENROLL(2,I)
      BPROJ(M) = BPROJ(M) - V4 * DINSTI(I,1) * ENROLL(2,I-1)
      IF (IGRADE .EQ. 13 .AND. I .GT. 8) GO TO 333
      IF (IGRADE .EQ. 12 .AND. I .GT. 7) GO TO 333
      GO TO 3333
  333 BPROJ(M) = BPROJ(M) - V5 *  DDROPS(L,1) * ENROLL(2,I-1)
 3333 BPRESY(I,M) = BPROJ(M)
30003 CONTINUE
      IF (ISEX .EQ. 0) GO TO 98765
      DO 70077 M = 1,100
      GPROJ(M) = TPROJ(M) - BPROJ(M)
70077 CONTINUE
98765 ICOUNT = 1
      CALL OUTPUT(I,J,TPROJ.ICOUNT,IGRADE,IDATE)
```

Program MAIN (continued)

```
0297          IF (ISEX .EQ. 0) GO TO 22222
0298          ICOUNT = 2
0299          CALL OUTPUT (I,J,BPROJ,ICOUNT,IGRADE,IDATE)
0300          ICOUNT = 3
0301          CALL OUTPUT(I,J,GPROJ,ICOUNT,IGRADE,IDATE)
0302          GO TO 22222
        C     THIS PART OF PROGRAM PROJECTS ENROLLMENT FOR SECOND (OR FIRST)
        C      GRADE THROUGH GRADE 12 (OR 13) FOR THE REMAINING YEARS OF THE
        C      SIMULATION
0303    83888 DO 40004 M = 1,100
0304          TPROJ(M) = TPREVY(I-1,M)
0305          CALL GAUSS(IX,SMIGRA(I,J),XMIGRA(I,J),V1)
0306          TPROJ(M) = TPROJ(M) + V1
0307          CALL GAUSS (IX,STRANS(I,J),XTRANS(I,J),V2)
0308          TPROJ(M) = TPROJ(M) + V2
0309          TPROJ(M) = TPROJ(M) - RHOLDS(M) * TPREVY(I-1,M)
0310          CALL GAUSS (IX,SHOLDS(I,J),XHOLDS(I,J),V3)
0311          RHOLDS(M) = V3
0312          TPROJ(M) = TPROJ(M) + V3 * TPREVY(I,M)
0313          CALL GAUSS(IX,SINSTI(I,J), XINSTI(I,J),V4)
0314          TPROJ(M) = TPROJ(M) - V4 * TPREVY(I-1,M)
0315          IF (IGRADE .EQ. 13  .AND. I .GT. 8) GO TO 34
0316          IF (IGRADE .EQ. 12 .AND. I .GT. 7) GO TO 34
0317          GO TO 205
0318       34 L = I - IGRADE + 5
0319          CALL GAUSS (IX,SDROPS(L,J),XDROPS(L,J),V5)
0320          TPROJ(M) = TPROJ(M) - V5 * TPREVY(I-1,M)
0321      205 TPRESY(I,M) = TPROJ(M)
0322          IF (ISEX .EQ. 0) GO TO 40004
0323          BPROJ(M) = BPREVY(I-1,M)
0324          BPROJ(M) = BPROJ(M) + V1 * DMIGRA(I,J)
0325          BPROJ(M) = BPROJ(M) + V2 * DTRANS(I,J)
0326          BPROJ(M) = BPROJ(M) - RHOLDS(M) * DHOLDS(I-1,J) *
             2BPREVY(I-1,M)
0327          BPROJ(M) = BPROJ(M) + V3 * DHOLDS(I,J) * BPREVY(I,M)
0328          BPROJ(M) = BPROJ(M) - V4 * DINSTI(I,J) * BPREVY(I-1,M)
0329          IF (IGRADE .EQ. 13 .AND. I .GT. 8) GO TO 44
0330          IF (IGRADE .EQ. 12 .AND. I .GT. 7) GO TO 44
0331          GO TO 4444
0332       44 BPROJ(M) = BPROJ(M) - V5 * DDROPS(L,J) * BPREVY(I-1,M)
0333     4444 BPRESY(I,M) = BPROJ(M)
0334    40004 CONTINUE
0335          IF (ISEX .EQ. 0) GO TO 98989
0336          DO 77007 M = 1,100
0337          GPROJ(M) = TPROJ(M) - BPROJ(M)
0338    77007 CONTINUE
0339    98989 ICOUNT = 1
0340          CALL OUTPUT (I,J,TPROJ,ICOUNT,IGRADE,IDATE)
```

Program MAIN (continued)

```
      IF (ISEX .EQ. 0) GO TO 22222
      ICOUNT = 2
      CALL OUTPUT (I,J,BPROJ,ICOUNT,IGRADE,IDATE)
      ICOUNT = 3
      CALL OUTPUT (I,J,GPROJ,ICOUNT,IGRADE,IDATE)
22222 CONTINUE
      DO 55555 M = 1,100
      DO 55555 I = 1,IGRADE
      TPREVY(I,M) = TPRESY(I,M)
      BPREVY(I,M) = BPRESY(I,M)
55555 CONTINUE
11111 CONTINUE
      STOP
      END
```

APPENDIX B

Subroutine OUTPUT

```
0001          SUBROUTINE OUTPUT(I,J,PROJ,ICOUNT,IGRADE,IDATE)
0002          DIMENSION PROJ(100), ENPROJ(11), PROB(11)
        C     ORDERING ROUTINE
0003        2 ISWCH = 0
0004          DO 1 M = 1,99
0005          IF (PROJ(M) .LE. PROJ(M + 1)) GO TO 1
0006          FINTER = PROJ(M)
0007          PROJ(M) = PROJ(M + 1)
0008          PROJ(M + 1) = FINTER
0009          ISWCH = 1
0010        1 CONTINUE
0011          IF (ISWCH .EQ. 1) GO TO 2
        C     ROUTINE FOR COMPUTING PROBABILITY LEVELS
0012          ENPROJ(1) = (PROJ(5) + PROJ(6))/2.
0013          ENPROJ(2) = (PROJ(10) + PROJ(11))/2.
0014          ENPROJ(3) = (PROJ(20) + PROJ(21))/2.
0015          ENPROJ(4) = (PROJ(30) + PROJ(31))/2.
0016          ENPROJ(5) = (PROJ(40) + PROJ(41))/2.
0017          ENPROJ(6) = (PROJ(50) + PROJ(51))/2.
0018          ENPROJ(7) = (PROJ(60) + PROJ(61))/2.
0019          ENPROJ(8) = (PROJ(70) + PROJ(71))/2.
0020          ENPROJ(9) = (PROJ(80) + PROJ(81))/2.
0021          ENPROJ(10) = (PROJ(90) + PROJ(91))/2.
0022          ENPROJ(11) = (PROJ(95) + PROJ(96))/2.
0023          PROB(1) = .05
0024          PROB(2) = .10
0025          PROB(3) = .20
0026          PROB(4) = .30
0027          PROB(5) = .40
0028          PROB(6) = .50
0029          PROB(7) = .60
0030          PROB(8) = .70
0031          PROB(9) = .80
0032          PROB(10) = .90
0033          PROB(11) = .95
0034          IF (IGRADE .EQ. 12) II = I
0035          IF (IGRADE .EQ. 13) II = I - 1
0036          JJ = J + IDATE - 1
        C     PRINTING ROUTINE
0037          IF (ICOUNT .EQ. 1) WRITE(6,7001) II,JJ,II,JJ
0038          IF (ICOUNT .EQ. 2) WRITE(6,7002) II,JJ,II,JJ
0039          IF (ICOUNT .EQ. 3) WRITE(6,7003) II,JJ,II,JJ
0040     7001 FORMAT (/////1H0,'PROBABILITY THAT TOTAL ENROLLMENT IN GRADE',1X,
            2I2,1X,'IN'1X,I4,14X,'PROBABILITY THAT TOTAL ENROLLMENT IN GRADE',
            31X,I2,1X,'IN',1X,I4/1H ,'WILL BE LESS THAN THE SPECIFIED PREDICTED
            4 ENROLLMENT',14X,'WILL BE GREATER THAN THE SPECIFIED PREDICTED ENR
            5OLLMENT'//1H ,4X,'PROBABILITY',5X,'PREDICTED ENROLLMENT',30X,'PROB
            6ABILITY',5X,'PREDICTED ENROLLMENT'//)
```

Subroutine OUTPUT (continued)

```
0041        7002 FORMAT (/////1H0,'PROBABILITY THAT MALE ENROLLMENT IN GRADE',1X,
                 2I2,1X,'IN',1X,I4,15X,'PROBABILITY THAT MALE ENROLLMENT IN GRADE',
                 31X,I2,1X,'IN',1X,I4/1H ,'WILL BE LESS THAN THE SPECIFIED PREDICTED
                 4 ENROLLMENT',14X,'WILL BE GREATER THAN THE SPECIFIED PREDICTED ENR
                 5OLLMENT'//1H ,4X,'PROBABILITY',5X,'PREDICTED ENROLLMENT',30X,'PROB
                 6ABILITY',5X,'PREDICTED ENROLLMENT'//)
0042        7003 FORMAT (/////1H0,'PROBABILITY THAT FEMALE ENROLLMENT IN GRADE',
                 21X,I2,1X,'IN',1X,I4,13X,'PROBABILITY THAT FEMALE ENROLLMENT IN GRA
                 3DE',1X,I2,1X,'IN',1X,I4/1H ,'WILL BE LESS THAN THE SPECIFIED PREDI
                 4CTED ENROLLMENT',14X,'WILL BE GREATER THAN THE SPECIFIED PREDICTED
                 5 ENROLLMENT'//1H ,4X,'PROBABILITY',5X,'PREDICTED ENROLLMENT',30X,
                 6'PROBABILITY',5X,'PREDICTED ENROLLMENT'//)
0043             WRITE (6,7000) PROB(1), ENPROJ(1), PROB(1),
                 2ENPROJ(11), PROB(2), ENPROJ(2), PROB(2), ENPROJ(10), PROB(3),
                 3ENPROJ(3), PROB(3), ENPROJ(9), PROB(4), ENPROJ(4), PROB(4),
                 4ENPROJ(8), PROB(5), ENPROJ(5), PROB(5), ENPROJ(7), PROB(6),
                 5ENPROJ(6), PROB(6), ENPROJ(6), PROB(7), ENPROJ(7), PROB(7),
                 6ENPROJ(5), PROB(8), ENPROJ(8), PROB(8), ENPROJ(4), PROB(9),
                 7ENPROJ(9), PROB(9), ENPROJ(3), PROB(10), ENPROJ(10), PROB(10),
                 8ENPROJ(2), PROB(11), ENPROJ(11), PROB(11), ENPROJ(1)
0044        7000 FORMAT (1H ,8X,F3.2,11X,F10.0,42X,F3.2,11X,F10.0)
0045             WRITE (7,5000) (PROJ(LJK), LJK = 1,100)
0046        5000 FORMAT (16F5.0)
0047             RETURN
0048             END
```

APPENDIX C

INSTRUCTIONS TO THE USER

The input deck should consist of the following cards:

I Title card
Columns 1-80 may be used. The title will be used as a heading on the output.

II Parameter card

VARIABLE		FORM	COLUMN(S)
IYEAR	(number of years to be simulated)	any integer from one to fifteen	9-10
IGRADE	(number of grades to be simulated)	either 12 or 13	14-15
ISEX	(to indicate use of sex option)	1 if yes, 0 if no	20
IDATE	(first year to be simulated)	19xx	22-25
IX	(number for random number generator)	any 9 digit odd integer	27-35

III Format cards for input parameters and variables.
Use a separate card for each of the 13 formats. Columns 1-80 may be used. The following formats are suggested, but not required.

Variable (V) or Parameter (P)		Format When Sex Option Is Not Used	Format When Sex Option Is Used
(P1)	ENROLL	(6(F5.0,5x))	(12F5.0)
(P2)	BIRTHS	(5(F5.0,5x))	(10F5.0)
(P3)	DEATHS	(5(F5.0,5x))	(10F5.0)
(V1)	ABIRTH, BBIRTH, CBIRTH	(3F4.0)	(3F4.0,4x,F4.3)
(V2)	AGIRLS, BGIRLS, CGIRLS	(3F5.0)	(3F5.0)
(V3)	ADEATH, BDEATH, CDEATH	(3F5.3)	(3F5.3,1x,F4.3)
(V4)	APREMI, BPREMI, CPREMI	(3F5.0)	(3F5.0,1x,F4.3)
(V5)	AMIGRA, BMIGRA, CMIGRA	(3F3.0)	(3F3.0,7x,F4.3)
(V6)	APRIVS, BPRIVS, CPRIVS	(3F4.0)	(3F4.0,4x,F4.3)
(V7)	ATRANS, BTRANS, CTRANS	(3F5.0)	(3F5.0,1x,F4.3)
(V8)	AHOLDS, BHOLDS, CHOLDS	(3F3.2)	(3F3.2,7x,F4.3)
(V9)	AINSTI, BINSTI, CINSTI	(3F4.2)	(3F4.2,4x,F4.3)
(V10)	ADROPS, BDROPS, CDROPS	(3F3.2)	(3F3.2,7x,F4.3)

INSTRUCTIONS TO THE USER (continued)

IV Following the format cards should be input parameter and variable data cards. These cards should follow the formats on the format cards. The following is a brief explanation of the parameters and variables. In the explanation, "base year" refers to the year immediately preceding the first year of prediction; often it is the year in which the prediction study is conducted. "Years of simulation" do not include the base year. The "previous year" refers to the twelve months previous to the simulation or prediction date, a date during the fall semester for which the predictions are made. The "year and grade of simulation" are the year and grade for which the predictions are being made.

PARAMETERS

(1) ENROLL (I) I = grade level	The number of children in each grade of the public schools at the beginning of school in the base year.
(2) BIRTHS (J) J = year of simulation	The number of allocated births in each of the four or five years previous to the base year, depending on whether kindergarten or first grade is the first level to be predicted. This assumes figures for the base year are not available and must be predicted.
(3) DEATHS (J) J = year of simulation	The number of deaths for these births.

VARIABLES

Each of the following variables has three or four forms, all of which must be included in the data. The various forms are designated by the prefixes "A," "B," "C," and "D." For example, ABIRTH refers to the "high" estimate; BBIRTH to the "most likely;" and CBIRTH, the "low" estimate. The optional form, DBIRTH, refers to the estimated proportion made for the variable. Other variables follow the same pattern.

Data should be ordered so that all forms of the variable with the first subscript values are read; then the three or four forms with the second subscript values are read, and so on. When two subscripts are used, the first subscript is the first to vary, forming the "inside" loop.

INSTRUCTIONS TO THE USER (continued)

(1)	ABIRTH (I, J) I = age group J = year of simulation minus four or five	The number of live births for each 1000 females in each of six age groups (15-19, 20-24, 25-29, 30-34, 35-39, and 40-44). Values for this variable will be required for the base year of simulation and all other years except the last four or five, depending on whether kindergarten or first grade is the first grade to be predicted.
(2)	AGIRLS (I, J) I = age group J = year of simulation minus four or five	The number of females in each of the six age groups for the years specified for Variable 1.
(3)	ADEATH (J) J = year of simulation	The proportions of preschool deaths for the births estimated by Variables 1 and 2.
(4)	APREMI (J) J = year of simulation	The net preschool migration in numbers of children migrating between birth and age of entry into kindergarten or first grade each year of simulation.
(5)	AMIGRA (I, J) I = grade of simulation J = year of simulation	For each year and grade of simulation, except the first grade level, net migration to the public schools during the previous year and grade. If there is a net gain, the value will be positive; if there is a net loss, it will be negative.
(6)	APRIVS (J) J = year of simulation	The number of potential kindergarten children or first graders who will enroll in non-public school instead of public school during each year of simulation.
(7)	ATRANS (I, J) I = grade of simulation J = year of simulation	For each year and grade of the simulation, net transfers to non-public schools during the previous year and grade. A loss in the public schools would be reflected by a negative net transfer and a gain reflected by a positive figure.

INSTRUCTIONS TO THE USER (continued)

(8)	AHOLDS (I, J) I = grade in which students remain J = year of simulation	For each year of simulation, the proportions of students in each grade who are retained at the end of the previous year and will remain in that grade level for another year.
(9)	AINSTI (I, J) I = grade of simulation J = year of simulation	For each year and grade of simulation, the percentage of students who are dropped from the rolls because of death or institutionalization during the previous year and grade.
(10)	ADROPS (I, J) I = grade of drop-out J = year of simulation	For each year of simulation, the proportion of students who dropped out of school from grades seven through eleven during the previous year.