DOCUMENT RESUME

ED 083 806 EM 011 516

AUTHOR	Zagorski, Henry J.; And Others
TITLE	Automatic Data Processing System and Procedures, Computerized Academic Counseling System.
INSTITUTION	System Development Corp., Santa Monica, Calif.
SPONS AGENCY	Air Force Human Resources Lab., Brooks AFB, Texas.
REPORT NO	AFHRL-TR-73-6
PUB DATE	Jun 73
NOTE	55p.
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	*Career Planning; *Computer Oriented Programs; Computer Programs; Counseling; Counseling Programs; *Electronic Data Processing; Guidance; Guidance Counseling; Military Personnel; Program Descriptions; Technical Reports; Undergraduate Study; *Vocational Counseling
IDENTIFIERS	CACS; *Computerized Academic Counseling System; System Development Corporation; United States Air Force

ABSTRACT

The Computerized Academic Counseling System (CACS) designed by the System Development Corporation is reviewed. Aspects of the system, constructed to assist counselors in guiding undergraduates in the selection of academic majors, which are discussed include: problem definition, system analysis, design rationale, methodology, measurement specifications, data base compilation, mathematical modeling, statistical results, and validation tests. Counseling application directions and capabilities are considered, computerized academic counseling in the context of career success likelihood is analyzed, and recommendations for extending the approach to additional aspects of career guidance are made. A concept for an Air Force (AF) career counseling system which permits individuals to shape their careers is developed. Its functional components include. a) an AF personnel needs and resources forecast model and, b) an AF mechanism which permits personnel to select and assure careers of their choice. Preliminary analyses are offered which indicate such a system is feasible and could have a significant impact on AF enlistment and turnover rates. Finally, recommendations are presented suggesting future research and development. (Author)

AFHRL-TR-73-6

AIR FORCE HUMAN RESOURCES LABORATORY

AUTOMATIC DATA PROCESSING SYSTEM AND PROCEDURES COMPUTERIZED ACADEMIC COUNSELING SYSTEM

By

Henry J. Zagorski
Lemont E. Southworth
Stuart E. Charlston
Gloria L. Grace
Russell L. Smith
System Development Corporation
Santa Monica, California 90406

AIR FORCE HUMAN RESOURCES LABORATORY
Brooks Air Force Base, Texas 78235

June 1973

Approved for public release; distribution unlimited.

AIR FORCE SYSTEMS COMMAND
BROOKS AIR FORCE BASE, TEXAS 78235

NOTICE

When US Government drawings, specifications, or other data are used for any purpose other than a definitely related Government procurement operation, the Government thereby incurs no responsibility nor any obligation whatsoever, and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data is not to be regarded by implication or otherwise, as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

AFHRL-TR-73-6

June 1973

AUTOMATIC DATA PROCESSING SYSTEM AND PROCEDURES
COMPUTERIZED ACADEMIC COUNSELING SYSTEM

By

Henry J. Zagorski
Lemont E. Southworth
Stuart E. Charlston
Gloria L. Grace
Russell L. Smith
System Development Corporation
Santa Monica, California 90406

Approved for public release; distribution unlimited.

AIR FORCE HUMAN RESOURCES LABORATORY
AIR FORCE SYSTEMS COMMAND
Brooks Air Force Base, Texas 78235

FOREWORD

The research effort was conducted under work unit ILIR-00-31, Development of a Computer-Based Model for an Advanced Academic Counseling System. The effort was supported by program element 61101F, Laboratory Director's Funds. This report covers part of the research conducted by Systems Development Corporation, Santa Monica, California, under Contract Number F41609-71-C-0028. Major James E. Wade acted as contract monitor for the Air Force Human Resources Laboratory.

This report has been reviewed and is approved.

HAROLD E. FISCHER, Colonel, USAF
Commander

ABSTRACT

The Computerized Academic Counseling System (CACS) was developed as a counseling aid. The system was designed to provide in depth analysis and forecasting of student performance as an aid to counselors in assisting students in the selection of academic majors in which they are most likely to succeed.

CACS was specially geared to meet the requirements of an advanced program of personnel management. It brings valid and comprehensive data analysis capability into the academic counselor's hands in a timely and efficient manner. It assists him in performing the guidance function and in conducting the research needed to advance knowledge in vital areas of human resource development. It provides:

1. Timely access to a comprehensive array of counseling information
2. User-Oriented interface procedures
3. Readily interpretable displays
4. Flexibility of operation and maintenance
5. Modular expansion capability

TABLE OF CONTENTS

Table of Contents (Continued)

LIST OF TABLES

List of Tables (Continued)

AUTOMATIC DATA PROCESSING SYSTEM AND PROCEDURES COMPUTERIZED ACADEMIC COUNSELING SYSTEM

I. INTRODUCTION

General

This report describes the design, development and evaluation of the Computerized Academic Counseling System (CACS). Specifically designed as a counseling tool, for college level, CACS consists, in part, of a multiple linear regression equation whose outputs are:

1. The predicted likelihood of success for a college student in any or all of the major fields of study;

2. a predicted point estimate of a student's Grade Point Average (GPA) in each major; and

3. a standard error of estimate for each predicted point estimate.

In addition, CACS provides, on request, a list of predictor values that are used in performing the GPA and probability of success calculations, and a summary of the student's academic grades.

CACS is a modularized system written in COBOL for use on a Burroughs B-3500 computer. The purpose of modularizing the program and writing it in COBOL was to (*a*) enable the system to be easily modified as curricula and majors change, and (*b*) permit adaptation of the system to operate on a variety of different computers with a minimum of reprogramming. In effect, CACS is a versatile model that is easy to use, and capable of modification and expansion; it will provide information that can be used by a counselor to successfully guide a student.

Problem Statement

It is generally agreed that not all high school graduates have the intellectual capability to succeed in college. Moreover, regardless of the number which might be viewed as potentially successful college students, available instructor staff and facilities pose severe constraints on the number of applicants that can be accepted by colleges. As a consequence, colleges generally accept only those high school graduates who have established a high grade point average. Thus, it is (at least implicitly) assumed that prior academic history is an essential and perhaps sufficient condition for predicting success in college.

High school students, junior college graduates; etc., who are precluded from attending college because of low academic grades, often are permitted a second opportunity, contingent on the results of a college administered battery of tests. Such tests usually consist of paced, simple arithmetic problems, vocabulary evaluations, verbal comprehansion and reasoning problems, and determinations of aptitude. Test batteries are validated and a cutoff line is drawn dichotomizing overall criteria scores into those which are acceptable and those which are unacceptable; it is assumed with a reasonable degree of confidence that students whose scores are below the cutoff line will not succeed in college.

Despite considerable efforts to develop methodologies which tend to ensure that accepted applicants will succed in college, the combined flunk-out and drop-out rate is relatively high, reaching as much as 50 percent in some universities.

At least three major factors can be delineated as contributing to student failure. First, many students enter these institutions with little or no consideration given to establishing academic (and future professional) goals or interests. As a result, such students perform poorly, change their major (and curriculum) frequently, and eventually flunk out or drop out. Second, many students simply demonstrate little motivation to study and their low grades provide cause for early termination. Finally, it is well recognized that the above noted selection process is fallible, permitting some applicants to filter through who are not, in fact, suited for the university environment because of a host of reasons, not the least of which is a specific intellectual capability to master the academic requirements.

Objective

Two implications can be drawn from the problem statement described above. First, prior academic history and test batteries provide reasonably good predictors of achievement in college, albeit with less accuracy than desired but far greater than that achievable by other techniques. Second, prediction success may increase dramatically if the "evaluation instrument" specifically takes into account student aptitudes and other factors related to curricula, thus maximizing the likelihood that students will select the correct curriculum.

In essence, the System Development Corporation (SDC) sought to develop a predictor instrument that could be used by counselors at any time to predict a student's likelihood of success in any of the major fields of study offered. Such an instrument would serve as an aid to counselors during the process of advising and counseling college students with respect to selecting the most appropriate major fields of study.

Design Rationale

General Design Criteria

The primary purpose of developing CACS was to provide counselors with a *tool* for facilitating the achievement of a basic counseling objective; namely, the selection of the most appropriate major for each student. Emphasis is placed on the term *tool* because the current state-of-the-art in predicting academic performance by mathematical modeling techniques is insufficient by itself, and we believe that the expertise and experiences of academic counselors should comprise the major influence on students seeking guidance. However, since academic performance is a function of many interrelated variables, the complexities of which cannot be precisely calculated and resolved by "pure" counseling, any objective tool that can reduce such complexities should facilitate the counseling process. The overall design criterion used in the development of CACS, therefore, was that it be a facilitator in the counseling process.

To be generally useful, the design of CACS would also have to comply with several specific and interrelated criteria. Such criteria are briefly discussed in the following sections.

Specific Design Criteria

1. *Long life expectancy*. The time and costs of developing a model, such as CACS, necessitates that it be designed for a relatively long life span. Thus, it was necessary to build into the model a modification capability in anticipation of wide variety of possible changes.

2. *Ease of expansion*. CACS was designed to account for current, available input data on each student; however, the model was also designed to accommodate additional inputs that would likely become available over time. Furthermore, a modularized design approach was used to add model flexibility, thus, modules can be eliminated, added or modified, as required.

3. *Ease of updating*. Long life span and flexibility gain increasing importance as the ease of updating the model's program increases. CACS would unlikely survive more than one or two updates, if the reprogramming tasks were unwieldly and difficult. Since many types of changes are expected to occur on a relatively frequent basis (*e.g*, modifications of majors, changes in formulae for computing GPA), it is apparent that simplicity in modifying the program was an essential design criterion.

4. *Simplicity of use*. The achievement of the above design criteria would be offset if model usage was too complex or time consuming. Thus. CACS was designed so that it can be readily employed with little effort on the part of the counselor or other college staff personnel.

5. *Timely response*. A final design criterion was the requirement for timely response. Clearly, the value of a tool is questionable if its availability is low when it is needed. Accessibility to all pertinent information should consume no more time than that normally occurring between the establishment of a counseling appointment and the actual counseling. CACS was, therefore, designed to provide, at maximum, an overnight response time.

Representative Applications And Examples

CACS is primarily intendent to serve as a tool for the academic counselor who is advising a student with regard to the selection of an academic major. However, it can also provide diagnostic information for the analysis of student academic problems and data for institutional research. Samples of each of these three immediate CACS applications (academic counseling, diagnostic evaluation and institutional research) are given in the examples below. Additional applications for CACS are sure to evolve once the system becomes fully operational.

Academic Counseling Application

For this example, let us consider the following counseling situation. A student is in his third class year and has not yet selected an academic major. He contacts his counselor for assistance in making his selection. If an appointment is made in advance, the counselor may request data via CACS batch processing capability for review before talking with the student. Or, if the student did not make an advance appointment, or the counselor does not wish to review the student's data prior to the counseling session, he may request data directly from the system via his terminal during the counseling session. At the terminal the counselor has the option of either CRT (television screen display) or hard copy (teletype) output—or both. The printouts or displays the counselor receives in response to either batch or on-line request are identical in content and format. The following is an example of a typical on-line session where the counselor enters his request via the terminal.

1. *Predicted GPA option.* Suppose student number 741234, John D. Brown, has requested counseling and has indicated he would like to major in mathematics. To determine his predicted GPA and probability of success in mathematics, the counselor enters the following request at his terminal:

GPAS 741234 MATH

CACS responds with the predicted GPA, standard error, and probability of success.

Example:

BROWN J D			741234
MAJOR	EST GPA	STD ERROR	PROB SUCCESS
MATH	2.55	.39	.92

A second alternative for obtaining this data is to request this information in all 28 majors by entering the following request:

GPAS 741234

CACS response to this request is a list of predicted GPA's in all 28 majors.

Example:

BROWN J D			741234
MAJOR	EST GPA	STD ERROR	PROB SUCCESS
AERO	2.80	.30	.99
AMERSTU	2.64	.31	.98
ASTRO	2.95	.27	.99
BASSCI	2.63	.42	.94
CHEM	2.82	.43	.97
CIVENGR	3.09	.37	.99
COMPSCI	2.69	.42	.95
ECON	2.67	.33	.98
ELENGR	3.06	.35	.99
ENGRMGT	2.72	.33	.99
ENGRSCI	2.62	.30	.98
FAREAST	2.52	.37	.92
GENENGR	2.80	.36	.99
GENSTU	2.54	.30	.96
GEOG	2.70	.45	.94
HISTORY	2.71	.42	.96
HUM	2.88	.37	.99
INTAFF	2.38	.30	.90
LATAMER	2.71	.27	.99
LIFESCI	2.98	.31	.99
MATH	2.55	.39	.92
MECH	2.99	.32	.99
MILARTSC	2.95	.29	.99
POLSCI	2.78	.35	.99
PHYSICS	2.82	.39	.98
PSYCH	2.81	.38	.98
SOVSTU	2.61	.35	.96
WESTEUR	2.43	.34	.89

Two additional printouts are available to assist the counselor in advising students. These are the grade and predictor summaries.

2. *Grade summary option.* Grade summaries may be requested for a single department or for the student's entire history. For example, if the counselor wants to examine the grades in mathematics, he could enter the request:

GRAD 741234 MATH

CACS response to this request is a list of courses taken in the specified department.

Example:

BROWN J D			741234
COURSE	NO.	HOURS	GRADE
MATH	100	5.50	X
MATH	161	6.00*	B
MATH	162	7.50*	B
MATH	232	2.50*	C
MATH	260	3.00*	B
TOTAL HOURS		19.00 GPA	2.87

To request a grade summary for the entire academic history, the counselor would enter the request:

GRAD 741234

CACS response to this request is a list of all courses the student has taken since entering. (Only a few of the first and last courses are shown in the sample below.) The printout is ordered by department. Courses flagged with an asterisk (*) beside the number of hours are those used in computing the total hours for each department and the group total shown.

Example:

BROWN J D			741234
COURSE	NO.	HOURS	GRADE
CHEM	121	2.50*	A
CHEM	122	3.00*	A
TOTAL HOURS		5.50 GPA	4.00
:	:	:	:
PHYSICS	220	5.00*	B
TOTAL HOURS		5.00 GPA	3.00
POL SCI	211	2.50*	C
POL SCI	212	3.00*	C
TOTAL HOURS		5.50 GPA	2.00
PSYCH	100	2.50*	C
TOTAL HOURS		2.50 GPA	2.00
GROUP TOTAL HRS		141.25 GPA	2.63

If the counselor wishes to know what predictors are influencing the GPA in a given major, he can request a predictor summary for that major. The predictors in MATH for our sample student would be obtained by entering the following request:

PRED 741234 MATH

CACS response to this request is a list of predictor values for the specified student. Predictors flagged with an asterisk (*) are those used in making the prediction for the specified major (in this case, MATH).

Example:

BROWN J D	741234
GPA - MATH	2.76*
GPA - OTHER BASIC SCIENCES	3.23*
GPA - ENG SCIENCES	3.00*
GPA - HUMANITIES	2.38
GPA - SOCIAL STUDIES	2.00*
FALCON/SKELLY SCHOLARSHIP	0
TURNBACK INDICATOR	0
ESTIMATE AGE AT GRADUATION	22
PRIOR ACADEMIC ACHIEVEMENT	575*
VERBAL APTITUDE	466
ENGLISH COMPOSITION	485
COMPOSITE ENGLISH SCORE	0951
MATH APTITUDE	684
INTERMEDIATE/ADV MATH CODE	1
MATH ACHIEVEMENT	698*
COMPOSITE MATH SCORE	1382
ACADEMIC COMPOSITE	2908
PAE SCORE	490
ACTIVITIES - ATHLETIC	570
ACTIVITIES - NON-ATHLETIC	490
LEADERSHIP COMPOSITE	1550
WEIGHTED COMPOSITE	562
MEDICAL QUALIFICATION CODE	1
ACADEMY PREP SCH ATTENDED	0
OTHER PREP SCH ATTENDED	0
COLLEGE ATTENDED CODE	0

If a list of the predictors without regard to any specific major is desired, the counselor can input the request:

PRED 741234

In this case CACS will respond with a list of predictor values without any asterisks as in the example below.

Example:

BROWN J D	741234
GPA - MATH	2.76
GPA - OTHER BASIC SCIENCES	3.23
GPA - ENG SCIENCES	3.00
GPA - HUMANITIES	2.38
GPA - SOCIAL STUDIES	2.00
FALCON/SKELLY SCHOLARSHIP	0
TURNBACK INDICATOR	0
ESTIMATE AGE AT GRADUATION	22
PRIOR ACADEMIC ACHIEVEMENT	575
VERBAL APTITUDE	466
ENGLISH COMPOSITION	485
COMPOSITE ENGLISH SCORE	0951
MATH APTITUDE	684
INTERMEDIATE/ADV MATH CODE	1
MATH ACHIEVEMENT	698
COMPOSITE MATH SCORE	1382
ACADEMIC COMPOSITE	2908
PAE SCORE	490
ACTIVITIES - ATHLETIC	570
ACTIVITIES - NON-ATHLETIC	490
LEADERSHIP COMPOSITE	1550
WEIGHTED COMPOSITE	562
MEDICAL QUALIFICATION CODE	1
ACADEMY PREP SCH ATTENDED	0
OTHER PREP SCH ATTENDED	0
COLLEGE ATTENDED CODE	0

All of the information obtained in the on-line examples, above, could also be obtained by keypunching the requests and submitting the cards for batch processing. The resulting printouts would be identical to the examples above.

Diagnostic Evaluation Application

The CACS system can very readily be used as a diagnostic tool by academic counselors. Both the on-line and the batch mode of processing are available for this purpose. For example, consider the situation where the counselor receives a list of students who are experiencing academic difficulties in their major.

Let us assume a fictitious student in his junior year, R.J. Green, student number 741235, is having difficulty as an electrical engineering major. The counselor can examine his grades in electrical engineering courses by entering the following request:

Example:

GREEN R J			741235
COURSE	NO.	HOURS	GRADE
EL ENGR	333	2.50*	C
EL ENGR	334	3.00*	D
TOTAL HOURS		5.50 GPA	1.4

Since the number of specific courses in electrical engineering is small, the counselor may wish to review all the courses the student has taken since he entered the school. To obtain this information he enters the request:

GRAD 741235

The CACS response to this request is a list of all courses taken by the student summarized by departments.

Example:

GREEN R J			741235
COURSE	NO.	HOURS	GRADE
AERO	331	2.50*	D
AERO	332	3.00*	D
TOTAL HOURS		5.50 GPA	1.00
ASTRO	432	2.50*	C
TOTAL HOURS		2.50 GPA	2.00
CHEM	101	2.50*	C
CHEM	102	3.00*	C
TOTAL HOURS		5.50 GPA	2.00
⋮	⋮	⋮	
LIFE SCI	210	2.50*	B
TOTAL HOURS		2.50 GPA	3.00
⋮	⋮	⋮	
MIL TNG	115	0.50*	D
MIL TNG	116	0.50*	C
MIL TNG	220	1.00	N
MIL TNG	220	1.00*	B
MIL TNG	320	2.00*	C
MIL TNG	320	2.00	N
TOTAL HOURS		4.00 GPA	2.13
⋮	⋮	⋮	
SOC	304	0.50*	B
TOTAL HOURS		0.50 GPA	3.00
SPANISH	101	2.50*	C
SPANISH	102	3.00*	C
TOTAL HOURS		5.50 GPA	2.00
GROUP TOTAL HRS		128.75 GPA	2.06

The asterisks, which appear by the class hours in the display, indicate that course is used in the computation of the GPA.

CACS can again be utilized to assist counselors in discovering solutions to these problems and also to assist the counselor in planning a more fitting program for the student. Examine the case of the student majoring in electrical engineering, as mentioned above. He is in his 2nd class year and has a limited amount of time before graduation, therefore, some changes in his curriculum may be suggested to better facilitate his progress. Provided the student has sufficient time remaining before graduation and his schedule can be arranged to accommodate the necessary subjects, a change in academic major may be a potential solution. The counselor can examine the likelihood of success for the student in other majors by entering the request:

GPAS 741235

The CACS response to this request is a list of predicted GPAs and related standard error and probability of success for the specified student in each of the 28 majors.

Example:

GREEN R J			741235
MAJOR	EST GPA	STD ERROR	PROB SUCCESS
AERO	1.92	.30	.39
AMERSTU	2.50	.31	.95
ASTRO	1.58	.27	.06
BASSCI	2.02	.42	.52
CHEM	1.62	.43	.19
CIVENGR	2.23	.37	.73
COMPSCI	2.09	.42	.58
ECON	2.17	.33	.70
ELENGR	2.13	.35	.65
ENGRMGT	2.30	.33	.82
ENGRSCI	1.72	.30	.18
FAREAST	2.43	.37	.88
GENENGR	2.05	.36	.56
GENSTU	2.20	.30	.75
GEOG	2.27	.45	.72
HISTORY	2.56	.42	.91
HUM	2.63	.37	.96
INTAFF	2.23	.30	.78
LATAMER	2.61	.27	.99
LIFESCI	2.61	.31	.97
MATH	1.61	.39	.16
MECH	1.68	.32	.16
MILARTSC	2.46	.29	.95
POLSCI	2.40	.35	.87
PHYSICS	1.61	.39	.16
PSYCH	2.49	.38	.90
SOVSTU	2.51	.35	.92
WESTEUR	2.30	.34	.81

Now, armed with the student's academic summary and an estimate of his performance in each of the 28 majors, the counselor can analyze the potential solutions to the student's problem. For example, in the sample case, the student shows a fairly low probability of success of .65 in his current major of electrical engineering, but a fairly high probability of success (.97) in life sciences. Also, he received a B in his life science courses. Assuming he has sufficient time remaining and his schedule can be appropriately arranged to complete all requirements for a new major in life sciences, a change of majors may be a solution to the student's academic problems.

Institutional Research Application

CACS may be used as an institutional research tool to study trends and characteristics with respect to different types of students for the purpose of advancing an understanding of major problem areas that may be associated with academic counseling.

The counselor may pursue an investigation of this type individually, or he may collaborate with other counselors or assistants. The first step the counseling researcher undertakes is an outline of the research design applicable to the institutional problem. In this step, he executes a simple statement of the problem, the tentative explanations or hypotheses that may apply, and the general technique by which data would be collected and analyzed to support or reject given solutions, explanations, or guidelines.

The CACS system offers the counselor an excellent capability for doing research concerning student academic counseling. This capability provides him with:

1. Easy access to most of the relevant data concerning student performance,

2. Up-to-date comprehensive information for individual students or groups of students in which he may be interested,

3. Meaningful formats requiring a minimum of data search and conversion,

4. Flexibility as to type of data desired,

5. Rapid response time to allow timely inquiries,

6. Accuracy of data made possible by effective computerization.

CACS could be used as the primary data collection technique, although data from other sources may also be included. In pursuing the analysis, the researcher defines pertinent study groups by stipulating the student control numbers for each group in which he is interested and selecting CACS options that are appropriate for the study. In an exhaustive study all options may be selected, thereby, providing a complete printout that reflects, for the student, his current predicted major GPAs and success probabilities, course grades, and descriptor variables. The selected student control numbers and CACS options may be prepared in punch card form for processing by CACS in the batch mode.

Simple but pertinent statistical analyses of each study group are performed by accumulating the data printed out by CACS. For example, if the study concerns the relative mathematical aptitude of a given group, the predictor summary for each student in that group is scanned for the mathematical aptitude indicator which is tabulated or transferred to worksheets for accumulation into frequency distributions, averages, percents, etc. The statistics computed during any given time period for a study group are recorded and compared with similar statistics computed for the same kind of group during a subsequent time period in order to isolate and identify trends. The trends are charted graphically using the means, percents or other statistics computed from the worksheets within any given time period. Group-to-group comparisons may also be made to determine those characteristics that distinguish one student group from another. Simple graphs reflecting group differences can also be made.

By means of the procedures described here the counseling researcher can pursue the answers to questions such as:

1. What are the characteristics of military prep school students that distinguish them from other students and, thereby, may be especially relevant in the counseling of those with prep school backgrounds?

2. Are there significant characteristics concerning course-to-course grade differentials that are especially important in the counseling of minority groups?

3. What student descriptor variables are pertinent in the differentiation of students with unusual cultural backgrounds such as Vietnamese, Chinese, etc?

4. Does the forecasting of major GPA performance, for students with very high intelligence or enriched academic background, create special problems in counseling?

5. Are counseling and student forecasting procedures adequate to meet the needs of female students? Although there are no data in CACS at present concerning female students, such data will be amenable to early analysis as soon as it becomes available in the system.

II. METHOD

Introduction

In developing multiple linear regression equations, the developer usually selects and measures input variables, a process which entails the development of test batteries and/or other data collection instruments. In the present project, however, the input variables were those normally measured by the school. The project did not call for the derivation of additional inputs. Since a regression model is only as valid as its input data, the predictive accuracy of CACS was, therefore, totally dependent on the data provided by the school.

Available Data

The data provided by the school consisted of: (*a*) Student Master Tape File (SMTF), containing 96 items of information accumulated over the years 1966 through 1971, and (*b*) Personnel Record Change File (PRCF), containing all grades accumulated over the years 1966 through 1971. In addition, fractional student data from personality and interest tests were also provided by the administration. All data were associated with students who graduated from the school.

Screening the Data for Relevancy, Sufficiency, and Usability

Many items in the SMTF were not relevant to the construction of a regression model. Such items included "social security number," "advisor code," and "current parent name." These items were eliminated from the data pool, resulting in the selection of 21 of the 96 original items for inclusion in the model. Two additional items, student ID number and major number, were used as code numbers for categorizing all data related to a given student.

All grades in the PRCF were used in constructing the model. The grades were analyzed to derive 10 GPA variables and one criterion variable; *i.e.*, the major GPA score of the last two school years upon which the model was designed to predict.

Data, related to the Cattell and Edwards Inventories, could not be used in the model because they were insufficiently distributed, with respect to the variables available in the SMTF and PRCF, to allow their integration into a multiple regression analysis.

The variables selected and/or derived in the model are listed below. (For detailed discussion of the variables and associated computations, see the Appendix, Mathematical Model Description and Maintenance).

Criterion Variable

Major GPA last 2 years

PRCF Derived Variables

(1) GPA - Math first 2 years
(2) GPA - Basic sciences first 2 years
(3) GPA - Engineering sciences first 2 years
(4) GPA - Humanities first 2 years
(5) GPA - Social sciences first 2 years
(6) GPA - Math first year
(7) GPA - Basic sciences first year
(8) GPA - Engineering sciences first year
(9) GPA - Humanities first year
(10) GPA - Social sciences first year

SMTF Variables

(11) Falcon or Skelly scholarship
(12) Turnback indicator
(13) Estimated age at graduation
(14) Preacademic achievement
(15) Verbal aptitude
(16) English composition
(17) Total English
(18) Math aptitude
(19) Advanced mathematics taken
(20) Math achievement
(21) Total mathematics
(22) Academic composite
(23) Physical aptitude examination
(24) Athletic activities
(25) Nonathletic activities
(26) Leadership composite
(27) Weighted composite
(28) Medical qualification
(29) Military prep school
(30) Other prep school
(31) College attendance

Model Development

A computerized multiple regression analysis approach was used as the major technique for developing the equations required in the model to predict academic success. (A detailed description of model development is contained in the Appendix.) The validity of such a model is a direct function of the size and relevance of the data base made available for statistical analysis. The full range of variables available from SMTF and PRC data files were considered as independent variables for the model. As summarized in the objective paragraph, page 9, 31 such variables were considered as predictors for the criterion variable, which was the major GPA. In simple form, these 31 variables can be expressed as a vector (or column of data) that, when weighted appropriately by a parallel vector composed of regression coefficients and adjusted for a regression constant, will generate a point estimate (Y) with respect to the major GPA in a specified field; *e.g.*, chemistry. This process can be summarized in the expression $BX + C = Y$; wherein, the Bs and Cs are adjusted statistically by the multiple regression analysis technique for each major within each student Level (A, B ,or C), depending on the unique correlation of each independent variable with the major GPA in past samples of students who have a complete performance record; *i.e.*, those who have graduated.

Variables that exhibit characteristics that are redundant with other variables or that show poor partial correlations are kept in the model for future development purposes, but are assigned a null regression coefficient that effectively cancels out their effects until such time as it is desirable to weight them in the model.

During analysis, 28 different majors were observed in the data base. A separate expression, similar to that described above, was developed for each major by three different student maturity levels where A represented a student with over 45 credit hours, B a student with 15-45 credit hours, and C a student with less than 15 credit hours. In all, 84 multiple regression equations were obtained and installed in the CACS Prediction Module. Each equation was designed to provide a point estimate of major GPA performance for any student being counseled.

Along with the point estimate, the model provides a standard error of estimate that could be used by the counselor in assessing the relative degree of confidence he could place in the point estimate. Initially, the standard error term placed in the model was derived from the equation derivation sample by examining the deviations between what the equation said for the sample as compared to the actual major GPA. This error term may eventually be replaced by a refined standard error of forecasting term, obtained by similarly examining point estimate deviations but for a new sample of students. This potential refinement is described in the construction of prediction equations paragraph, page 14.

A third capability developed for the model was an algorithm for computing the probability that a student will be successful in a given field of academic endeavor. Here the point estimate of grade performance (previously described) is combined with the standard error term and a constant representing the minimum permissible major GPA performance level at the school (2.00) to provide the required probability statement. The procedure consisted of expressing the minimum acceptable performance as a deviation from the point estimate. This deviation is converted into a standardized Z-ratio expressed in units of the standard error of estimate. This ratio is converted via standard normal curve table transformation into a probability statement that reflects the extent to which the student's actual major GPA is likely to fall into an acceptable performance region.

In the process of selecting a prediction vehicle for forecasting student performance, five types of modeling approaches were examined.

1. Correlation-based models, including multiple regression analysis and the use of joint-occurrence matrices

2. Cluster-oriented models that make use of factor analytical techniques

3. Tree-structure models that treat data in sequential foliations

4. Pattern analysis models where subsets of similar performance characteristics are used to define properties of the model

5. Rational modeling where predictions are made on the basis of data-guided estimates made by the counselor.

While evaluation potential modeling approaches, it was also necessary to examine the amount and characteristics of the data available, the availability of central and peripheral computer hardware, and the school's overall software configuration. Criteria utilized included:

1. Parsimony of assumptions required for model employment

2. Demonstrated effectiveness in actual application

3. Efficiency in all modes of operation

4. Flexibility and adaptability

Cluster-oriented models would probably provide interesting aspects of student performance that could be expressed as major themes or factors in academic work, but they would not have solved the direct requirement for achieving a visible estimate of academic performance. Thus, they were eliminated from consideration in the initial development phase.

Decision tree structural models, where students are analyzed on branches of an analytical tree, provide yet another potential and interesting recourse to investigation of student data. However, they tend to duplicate the results of the multiple regression approach, while requiring considerable sophistication in data manipulation on the part of the analyst. They also tend to require elaborate computer capacity; thus, they were excluded from consideration for CACS.

Pattern analysis models were considered too difficult to achieve in the time allowed, and they also typically lack the visible forecasting capability provided by multiple regression analysis. The cost of these models is difficult to control and usually runs to excessive proportions.

Rational modeling will be used to some extent by all counselors regardless of the formal model used in the computer system. No formal attempt was made to structure this type of model since it was considered impractical to collect, specify, and program all the different possible configurations that different counselors employ in their characteristic approaches to student guidance.

Based on a rigorous examination of the five types of modeling approaches utilizing the criteria identified above, the multiple regression analysis approach was selected as the one that would (*a*) provide the greatest initial and long-term benefits, (*b*) be the easiest to install and maintain, (*c*) lend itself best to modification and expansion, and (*d*) be the most cost-effective.

Computer Program Development

CACS program modules were designed around the basic function of the system utilizing one module per function. The four modules and their respective functions are:

1. Control module

2. Data retrieval module

3. Prediction Module

4. Data base extraction module

This modular approach not only simplified program development and maintenance, but also makes use of the overlay capability of the B-3500 COBOL compiler to reduce the amount of computer storage required to operate the system.

Data that are likely to require frequent updates, such as predictor values, are allocated to external files that are easily modified by the data base extraction module. This module maintains a set of values that accurately describes the current enrollment at the school.

System commands have been kept simple and direct to enable a user to quickly learn the system. Outputs for display and hard copy terminals were designed to present data in a concise and meaningful way. All counseling aids commands in the system are available at the remote terminal and as a batch job via card input at the computer, to provide the user with optimum methods of retrieving data. For example, a counselor can request fairly large amounts of data via batch processing before a counseling appointment with a student, and/or retrieve additional information on-line from the computer via his remote terminal during the interview.

In general, CACS has been designed as a straightforward and usable tool for academic counselors. This report has deliverately avoided complicated system operation procedures, thereby, reducing training and maintenance efforts while increasing utility.

Model Capabilities

The capabilities of the four CACS program modules are presented below. (A detailed description of the modules is reserved in the project file for ILIR-00-31.)

Control Program Module

This program module, performing the executive function of the program system, ensures the parameters required for CACS interface with the school's Burroughs B-3500 computer system.

Prediction Program Module

This program module contains the equations necessary for computing achievement predictions for a given student. It represents the core of CACS and is designed to execute all forecast and summarizing computation and output resulting data. The prediction program module was designed to provide the three outputs noted in the general paragraph, page 1; *i.e.*, probability of success in major field, GPA point estimate, and a standard error associated with the point estimate. The three output predictions are determinable for each 28 major fields of study. Thus, a counselor can acquire a total predictive profile consisting of 28 GPA predictions, 28 standard errors, and 28 success probabilities on a given student for all curricula. If desired, a printout of all predictor values used in the derivation of the predictions and all academic grades can be executed.

Data Retrieval Program Module

The data retrieval program module retrieves the required student data from the CACS data base stored on the disk. This model operates in conjunction with the prediction program module and does not require user intervention.

Data Base Extraction Program Module

This program module is used in the maintenance and creation of the CACS data base which is subsequently stored onto disk for use by the CACS system.

Specific Applications

Three sets of equations were designed for each model--Level A, B, and C. Every effort was made to maximize the utility of each level. The project investigators recognized that the earlier in the academic career of a student that achievement predictions can be acquired, the greater the utility and effects of counseling guidance. It should be recognized that a concomitant of early application is decreased model validity, due to fewer input data and lesser quality of data. Such an outcome is inevitable by the very nature of prediction models and was an hypothesized expectancy. It is stressed here to ensure that the user be cognizant of the model's limitations in providing counseling guidance information.

Brief Summary of Project Developmental Tasks

The CACS project included six distinct, but related tasks:

1. Conduct an alternate system study
2. Design, develop, and test a prediction model
3. Design, develop, and test real-time computer programs
4. Procure, integrate, and deliver display and hardcopy equipment
5. Provide documentation and briefings
6. Deliver, install, and validate the system

Task 1 - Conduct An Alternative System Study

A complete description of Task 1 and resulting recommendations, is reserved in the Project file for ILIR-00-31 concerning an Analysis and Evaluation Alternative Computer System Configurations for the Computerized Academic Counseling System. This task involved a thorough description of system requirements, an analytical discussion of each of three alternative computer systems that could be used for CACS, cost comparisons of the three alternatives, and rationale for selecting a specific alternative.

Task 2 - Design, Develop, and Test A Prediction Model

Task 2 is thoroughly described in the appendix. This task involved the development of the selected mathematical model used in the derivation of prediction equations and of methodology required for updating these equations as additional data become available.

Tasks 1 and 2 were performed concurrently to provide timely design data for use in Tasks 3 and 4.

Task 3 - Design, Develop, and Test Real-Time Computer Programs

Task 3 is documented in detail in the project file for ILIR-00-31. It consisted of developing: (*a*) necessary computer programs (*b*) specifications for operating the system (c) step-by-step procedures for using CACS and (d) methodology for maintaining the integrity of the system.

The design philosophy employed emphasized a modular approach and the utilization of existing or standard computer facilities. This philosophy provides for the most efficient use of system resources and facilitates updating, modifying and expanding the system to keep pace with academic counseling requirements and changing computer facilities.

Task 4 - Procure, Integrate, and Deliver Display and Hardcopy Equipment

Task 4 involved the acquisition and integration of display and hardcopy equipment for two different terminal configurations to be used in four CACS counselor positions. Both configurations consist of a cathode ray tube display terminal, hardcopy printer and a data set. The same data sets are used in both configurations, Omnitec 701A Acoustic Telephone Couplers. These couplers are capable of operating over either leased or dialed lines and interface with both Teletype (TTY) and RS232 terminal devices.

The first CACS terminal configuration consists of a Data Point 3000 Display Terminal, a Teletype Model 33 RO (Read Only) for hardcopy output, an Omnitec 701A Acoustic Coupler. The second terminal configuration is composed of a Teletype Model 33KSR (keyboard send/receive), and Ann Arbor 202 display terminal with a 9-inch video monitor, and an Omnitec 701A Acoustic Coupler.

Each of the terminal configurations is capable of being operated over either leased or dialed telephone lines, and with or without, hardcopy output. System operations and test procedures for CACS are given in the project file for ILIR-00-31, CACS Installation and Test.

Task 5 - Provide Documentation and Briefings

This task is self-evident; the present technical report and documentation, cited elsewhere within this report, constitute the products of this task.

Task 6 - Deliver, Install, and Validate the System

Task 6 consisted of: (a) exercising the model using the criterion input GPA variable derived from the class of 1972 through the fall semester of 1971, (b) computing the multiple linear regression equation predictions, and (c) subsequently validating the derived equations. All subtasks have been completed.

III. RESULTS

Introduction

Data for 3234 graduated students concerning 31 potential predictor variables with regard to major GPA were extracted from the CAIDS-PRC files for the time period 1967 through 1971. Due to technical difficulties associated with different tape design, 1966 student information could not be incorporated into the basic data. A study of the sample sizes for the majors, involved in 1966, indicated that the omission of this data would not seriously disturb the equations to be derived. Punched card data, concerning Cattell and Edwards personality information for students during the time period 1968 through 1971, were received from the school, but these could not be properly integrated with the SMTF-PRC files for multiple regression analysis purposes.

Construction of Prediction Model

Data for 3234 graduated students were partitioned by major areas of study across graduating years as shown in Table 1. The graduates in each major were combined into sample study groups. Frequency distribution and variance analyses were conducted across all majors for the 31 numerical variables extracted from the SMTF-PRC files. These variables are summarized in the screening paragraph, page 8. The detailed logic for their derivation is explained in the appendix.

The variance ratios examined were statistically significant. This was particularly true for major GPA, indicating that inter-major differences were sufficient to preclude grouping of students into larger groups for the purpose of predicting grade performance. Therefore, separate multiple regression equation models for each major were derived. In addition, three levels of student maturity, at which the CACS system would operate in making grade point average forecasts, were established.

These levels were as follows:

Level A - the student has over 45 credit hours.

Level B - the student has 15-45 credit hours.

Table 1. Distribution of Graduated Students by Years and Majors

MAJOR	CODE	1967	1968	1969	1970	1971	TOTAL
HUMAN	01	10	8	5	5	2	30
BASSC	02	103	48	25	17	6	199
ENGSC	03	40	53	37	15	16	161
INTAF	04	87	92	55	74	50	358
MILSC	05	8	6	10	5	2	31
MATH	06	42	35	20	18	26	141
ASTRO	07	32	49	51	30	25	187
HIST	08	10	18	21	25	36	110
ENGMG	09	63	74	83	59	92	371
CVENG	10	24	34	47	40	34	179
ELENG	11	12	20	17	20	27	96
ENGMC	12	8	9	15	22	42	96
CHEM	13	13	9	11	9	7	49
PHYS	14	10	17	11	25	17	80
ARENG	15	14	22	50	60	47	193
PSYCH	16	8	19	17	13	16	73
ECON	17	12	14	31	31	23	111
POLSC	18	0	11	9	27	5	52
GEOG	19	0	2	14	12	7	35
AM STD	20	0	1	6	3	4	14
GNSTD	21	19	46	74	126	64	329
GNENG	22	2	15	16	12	27	72
CPTSC	23	0	0	23	24	42	89
LIFSC	24	0	0	0	33	49	82
FESTD	25	0	0	3	11	6	20
LASTD	26	0	0	11	16	12	39
SVSTD	27	0	0	2	7	9	18
WESTD	28	0	0	8	6	5	19
TOTAL		517	602	678	745	692	3234

Level C - the student has less than 15 credit hours.

Modeling by levels necessitated the construction of 84 different multiple regression equations for installation in the prediction module of the system.

Construction of Prediction Equations

Each of the 31 potential predictor variables was statistically evaluated for inclusion in each of the 84 equations. On the average, about 5 computer runs were required for each regression analysis or a total of about 400 runs. The detailed technical procedures, by which predictor variables were screened and selected, is described in the appendix. The analytical data and results used in equation development appear in the project file for ILIR-00-31.

Level A Model

The results of the Level A Model development are summarized in Table 2, which shows in column 2 the multiple correlation for each major. By squaring this value, the specific amount of statistical efficiency (100 X R^2) obtained in the data sample for each equation.

The statistical efficiencies range from 29 percent for Computer Sciences majors to 70 percent for American Studies majors. The efficiency for the latter group is believed to be inflated by the small sample size of 14 students on which that equation was based. However, the efficiences for Engineering Mechanics and International Affairs majors are quite high (64 percent, 62 percent) and are not considered to be inflated by small sample sizes. The efficiences for other majors are correspondingly high. The overall percent of performance variance accounted for by all equations at this level is 49 percent, which compares favorably with most similar studies of university performance. This supports the conclusion that the mathematical Level A Model is generally quite efficient in forecasting major GPA performance.

Table 2 also shows the relative percent statistical contribution of each selected predictor variable to the forecasting efficiency of each equation. The bottom row of Table 2 exhibits the summary impacts of each predictor across all majors. Variables are arranged from left to right in ascending order of impact on major GPA. Thus, it can be observed that the largest impact occurred for Social Sciences GPA (14 percent), while the least impact occurred for English Composition (.3 percent).

In summary, The Level A Model relies heavily on early GPA revealed during the first two years at the school but includes some weight (about 2 percent) for available pre-college variables. The Level A Model should reflect a high degree of utility for counselors, when they are working with students at the critical decision point in their academic life; *i.e.*, when students must choose a major field of endeavor.

Level B Model

The results of the Level B Model development are summarized in Table 3. Here, the statistical efficiencies range from 14 percent for Computer Sciences to 70 percent for American Studies. The efficiencies for the three top majors are probably inflated by small sample size and should be discounted. However, the efficiences for International Affairs (51 percent) and Mathematics (50 percent) are not inflated by small sample sizes and can be regarded as being highly respectable in the forecasting realm. The statistical efficiencies for the other majors are lower, but still considered quite respectable for prediction purposes. The overall efficiency of the Level B Model was 38 percent as compared to the 49 percent observed for Level A. This clearly indicates that the Level A Model is superior and should be used whenever possible to assess future student performance.

The availability of the Level B Model enhances the range of operational utility for CACS, since it can be applied earlier than the Level A Model. Considering that the model currently has no access to motivational factors, occupational interest, or specific aptitudes, the 38 percent overall efficiency appears to be quite acceptable for use in counseling at an intermediate point in the students academic life; *i.e.*, sometime prior to the time that a major must be selected.

As in the previous table, the selected predictor variables in Table 3 are arranged in ascending order of impact. For the Level B. Model, the largest impact across all majors occurred for Humanities GPA (10 percent) while the least occurred for Mathematical Aptitude (1 percent).

In summary, the Level B Model relies heavily on first year GPA and includes an increased weight

Table 2. Academic Performance Prediction Model
(Maturity Level A*)

MAJOR ABBREV.	R**	N	Pre-College Predictors: Engl Comp	Verb Apt	Math Acht	Math Apt	PRIOR ACAD	4th and 3rd Class Year Predictors: GPA MATH	GPA BAS SC	GPA ENG SC	GPA HUMAN	GPA SOC SC
AM STD	.84	14					34				27	9
ENG MECH	.80	96						11	27	18	5	3
INTAF	.79	358		3			4	2		3	19	31
MATH	.77	141			6		4	25	6	3		15
LA STD	.77	39					17				14	28
ASTRO	.76	187			1			13	15	29		
FE STD	.75	20					6				25	26
ARENG	.75	193						6	8	27	5	10
ENG SC	.73	161				1	3	13	7	21		8
ECON.	.73	111		2					15	4	9	23
ENG MGT	.71	371				1			6	7	6	31
LIF SC	.71	82	4						20			26
PSYCH	.70	73						17			13	19
GEOG	.69	35				5			6		12	25
POL SC	.69	52					6	4	9		11	17
MIL SC	.69	31					7	11		18		11
CIV ENG	.68	179						5	14	15	6	6
GEN ENG	.68	72				2	11	11	2	16		4
PHYSICS	.67	80			4			10	20	4	4	3
HUMAN	.66	30	3						2	6	26	7
EL ENG	.66	96				9			4	25		5
WE STD	.64	19					9				28	4
HIST	.63	110						5			21	14
CHEM	.62	49		5	4			2	25		3	
BAS SC	.61	199					8		12	4	9	4
GEN STD	.59	329					4		5	4	8	14
SV STD	.57	18					5				3	25
CPT SC	.54	89						3		17	4	5
AVERAGE MULTIPLE R	.69		.3	.4	.6	.7	4	5	7	8	9	14

*Level A Model is based primarily on predictor data compiled during students' first two years at the Academy during the time period 1967 through 1971.

**Represents the multiple correlation coefficient of the predictor variables with major GPA. R was squared and multiplied by 100 to obtain the figures discussed in the text.

Table 3. Academic Performance Prediction Model
(Maturity Level B*)

MAJOR ABBREV	R**	N	Pre-College Predictors: MATH APT	ENGL COMP	VERB APT	MATH ACHT	PRIOR ACAD	4th Class Year Predictors: GPA MATH	GPA SOC SC	GPA BAS SC	GPA HUMAN
AM STD	.84	14			14		33				23
LA STD	.77	39					12		23		25
FE STD	.77	20		7			5		31		16
INTAF	.71	358			5		7		6	10	23
MATH	.71	141				7	5	17	8	9	4
ENG MECH	.69	96					2	16	5	25	
LIF SC	.69	82		4				6	14	21	3
ECON	.66	111			5		3		11	12	13
GEOG	.65	35	4						6	22	10
POL SC	.64	52					6		7	15	13
ASTRO	.63	187				4		19	4	13	
ENG MGT	.62	371	2				1	2	7	7	19
WE STD	.62	19					12		2		24
ENG SC	.61	161	4				4	16	3	6	4
PSYCH	.60	73						17	5		14
ARENG	.59	193				3		11	3	10	6
MIL SC	.59	31					8	6		6	15
GN ENG	.58	72	3				16	15			
EL ENG	.57	96	12				3	8		10	
HIST	.57	110			1		1	5	11		15
CV ENG	.57	179						5	9	10	6
PHYSICS	.57	80				4		6	9	10	5
BAS SC	.54	199					10	1	1	9	6
SVSTD	.53	18		10			4				14
GNSTD	.51	329			2		6		4	4	10
HUMAN	.48	30		8					2	4	9
CHEM	.46	49		1	5	9		4	2		
CPT SC	.37	89						5			9
AVERAGE MULTIPLE R	.61		1	1	1	1	5	6	6	7	10

*Level B Model is based primarily on predictor data compiled during students' first year at the Academy during the time period 1967 through 1971.

**Represents the multiple correlation coefficient of the predictor variables with major GPA. R was squared and multiplied by 100 to obtain the figures discussed in the text.

(about 8 percent) for available pre-college predictors.

Level C Model

The results of the Level C Model development are summarized in Table 4. The statistical efficiencies for this model range from 2 percent for Computer Sciences to 62 percent for American Studies. The latter figure can be discounted due to the small sample on which it was based. The overall efficiency of the Level C Model was 19 percent, which was about 50 percent below that of Level B and about 60 percent below that of Level A. This is a considerable drop in statistical efficiency and some of the lower powered equations may be questioned, with respect to their validity in assessing student future performance at the time he first arrives at the school. However, considering the sharp gains available from the Level A and B Models and the generally better than chance efficiency of the Level C Model, the three models together provide good capability for use in progressive counseling.

The state-of-the-art in student career forecasting will be improved significantly as CACS is applied in counseling. There is little doubt that additional improvements can and will be possible after the system first becomes operational. These improvements will extend to the three maturity models described here as well as other models that can be incorporated into the approach. A discussion of several improvement possibilities is provided in the recommendations paragraph, page 26.

Prep School Effects on the Prediction Model

A special investigation of the 1967-1971 graduate data compiled during model development was conducted for the purpose of examining predictor-criterion differentials for students who attend prep schools.

The students were divided into five groups as follows:

1. No previous college or prep school (N=2360)
2. Military prep school (N=301)
3. Other prep school (N=65)
4. Previous college (N=457)
5. Not classifiable (N=51)

Within each group, grade performance was partitioned into the following subject matter areas:

1. Basic sciences (other than mathematics)
2. Social sciences
3. Engineering sciences
4. Humanities
5. Mathematics

Separate grade point average compilations were made for the freshmen and the freshmen plus the sophomores. The tabulations derived from this analysis are shown in Table 5, where it indicates that the military prep school attendees were clearly inferior in performance to the non-preps in every respect but one--freshman Mathematics. In this subject matter area, the military preps actually exceeded the non-preps; however, the difference was not statistically significant. The math GPAs for other prep school attendees did not stand out in this respect as much as those for military preps; however, they also are inflated in relation to the other subject matter GPAs.

Data for students who had no previous college prep school and those who attended military prep school were analyzed still further by partitioning among majors who had enough prep school attendees to allow meaningful comparisons to be made. The results of this analysis are shown in Table 6. The same general pattern as that observed in Table 1 prevails. For all but a few majors; *i.e.*, Civil Engineering, Computer Sciences, and Electrical Engineering, the preps surpass the non-preps in Mathematics during the freshman year. The difference between the preps and non-preps is substantially narrowed during the sophomore year. This can be observed in the mean differences in GPAs for the freshman and sophomore years, as compared to the sophomore year alone.

In every other subject matter area, the prep school graduates are clearly and consistently inferior to the non-prep graduates during the freshman and freshman plus sophomore years. This inferiority also extends to major GPA performance during the junior and senior class years.

The graduates who attended a previous college compare very favorably with those who had no prep school attendance. Consequently, these two types of students appear to be mergable for purposes of model derivation.

Table 4. Academic Performance Prediction Model
(Maturity Level C*)

			Pre-College Predictor						
MAJOR ABBREV.	R**	N	INTR/ADV MATH	EST. AGE GRADTN	MATH APT	ENGL COMP	MATH ACHT	VERB APT	PRIOR ACAD
AMSTD	.79	14		1		24		12	25
INTAF	.58	358		3		5		11	15
LASTD	.57	39						6	27
MATH	.54	141		1	2	1	13		12
FESTD	.54	20				5		7	17
ELENG	.50	96	3	1	15				6
GNENG	.50	72			4				21
ECON	.48	111						12	11
MIL SC	.48	31	3				4	5	11
CHEM	.46	49				2	13	6	
PHYSICS	.46	80	6	2			11		2
SVSTD	.42	18		6		6			6
WESTD	.42	19					12		6
BAS SC	.42	199		1				[illegible]	16
ENG SC	.40	161		2	5			2	7
GEOG	.40	35			6				10
ASTRO	.37	187		3	3		5	[illegible]	[illegible]
ENG MGT	.37	371			2		1	5	6
POL SC	.37	52				1		[illegible]	12
GN STD	.37	329					1	3	10
HIST	.36	110		2				6	5
HUMAN	.35	30		1		10		1	
LIF SC	.33	82		2		5		1	3
AR ENG	.32	193	1				5		4
CV ENG	.30	179		2	1				6
ENG MECH	.28	96						[illegible]	7
PSYCH	.20	73		2					2
CPT SC	.14	89					1		1
AVERAGE MULTIPLE R	.42		1	1	1	2	2	3	9

*Level C Model is based on pre-college data for students, collected during the time period 1967-1971.

**Represents the multiple correlation coefficient of the predictor variables with major GPA. R was squared and multiplied by 100 to obtain the figures discussed in the text.

Table 5. Mean Grade Point Averages for Different Types of Graduates During 1967-1971

Major	Class Years	No Previous College or Prep School (N=2360)	Military Prep School (N=301)	Other Prep School (N=65)	Previous College (N=457)	Grand Total (N=3183)
Basic Sciences	4th	2.91	2.48**	2.49**	2.91	2.86
(other than math)	4th,3rd	2.92	2.46**	2.54**	2.87	2.86
	4th	2.83	2.51**	2.60*	2.77	2.79
Social Sciences	4th,3rd	2.88	2.56**	2.60**	2.79**	2.83
Engineering Sciences	4th	3.09	2.71**	2.47**	3.02	3.02
	4th,3rd	2.90	2.59**	2.43**	2.79**	2.84
Humanities	4th	2.88	2.67**	2.70*	2.80*	2.85
	4th,3rd	2.88	2.64**	2.65**	2.78**	2.84
Mathematics	4th	2.89	2.92	2.70*	2.82*	2.88
	4th,3rd	2.89	2.85	2.66**	2.80**	2.87
All Majors	2nd,1st	3.01	2.76**	2.84*	3.01	2.98

*deviation from first column is significant at .05 level.
**deviation from first column is significant at .01 level.

NOTE: This table is arranged by subject matter categories and class year composites.

Table 6. Mean Grade Point Averages for Military Prep Students Compared to Non-Prep Students

MAJOR			MATH		HUMANITIES		ENGRG SCIENCES		SOCIAL SCIENCES		OTHER BASIC SCIENCES		MAJOR GPA
		N	4th	4th,3rd	4th	4th,3rd	4th	4th,3rd	4th	4th,3rd	4th	4th,3rd	2nd,1st
General Studies	NP	224	2.48	2.46	2.49	2.50	2.35	2.34	2.44	2.44	2.38	2.41	2.50
	AP	55	2.70	2.65	2.38	2.38	2.12*	2.24	2.22	2.17	2.11	2.10	2.37
Basic Science	NP	149	2.85	2.82	2.63	2.67	3.02	2.69	2.61	2.64	2.93	2.85	2.63
	AP	20	3.16	3.05	2.53	2.41	3.50*	2.55	2.37	2.36	2.62	2.64	2.46
General Engrg	NP	51	2.72	2.74	2.59	2.59	3.09	2.61	2.47	2.49	2.71	2.73	2.68
	AP	12	2.85	2.75	2.40	2.47	2.67*	2.51	2.37	2.40	2.44	2.47	2.56
Psychology	NP	51	2.57	2.56	2.72	2.77	3.50*	2.61	2.60	2.71	2.51	2.53	3.07
	AP	12	2.74	2.65	2.66	2.59	2.65*	2.66	2.49	2.41	2.23	2.11	2.88
Civil Engrg	NP	134	2.99	2.98	2.61	2.66	2.99	2.81	2.64	2.69	2.78	2.85	3.10
	AP	13	2.78	2.65	2.10	2.18	2.75*	2.85	2.19	2.12	2.59	2.43	2.93
Computer Science	NP	68	3.19	3.15	2.68	2.64	3.23	3.20	2.76	2.73	3.22	3.16	2.83
	AP	8	3.11*	3.08*	2.52*	2.52*	3.32*	3.29*	2.72*	2.67*	2.89*	2.98*	3.20
Engrg Science	NP	110	3.02	3.00	2.87	2.88	3.32	2.91	2.86	2.88	3.14	3.16	2.77
	AP	21	3.36	3.29	2.84	2.80	2.83*	2.76	2.78	2.84	2.87	2.93	2.68
Engrg Mgt	NP	263	2.66	2.66	2.86	2.83	2.76	2.82	2.74	2.89	2.70	2.71	3.01
	AP	39	2.86	2.81	2.68	2.72	2.57*	2.60	2.53	2.72	2.38	2.36	2.82
Intl Affairs	NP	265	2.64	2.66	3.17	3.14	2.68	2.75	3.02	3.04	2.74	2.72	3.08
	AP	37	2.76	2.91	3.14	2.98	2.50*	2.59	2.75	2.86	2.55	2.45	2.90
Elec Engrg	NP	68	3.24	3.22	2.91	2.86	3.43	3.29	2.96	3.01	3.30	3.38	3.30
	AP	12	2.84	2.80	2.81	2.78	3.00*	2.91	2.61	2.74	2.89	2.87	3.02
Aero Engrg	NP	146	3.21	3.19	2.89	2.92	3.21	3.10	2.86	2.93	3.26	3.29	3.06
	AP	10	3.34	3.23	2.74	2.83	3.00*	3.03	2.86	2.85	3.22	3.17	3.04
All Majors	NP	2360	2.89	2.89	2.88	2.88	3.09	2.90	2.83	2.88	2.91	2.92	3.01
	AP	301	2.92	2.85	2.67	2.64	2.71	2.59	2.51	2.56	2.48	2.46	2.76

*These means are based on fewer than 10 students.

NOTE: This table is arranged by subject matter categories and class year composites.

The conclusion derived from these comparisons was that the grade point average in the mathematics subject matter area for 4th class year students, who had attended the military prep school, were higher than those of other students with similar abilities. This elevated the grade point average in mathematics, though reduced during the sophomore year, appears to carry through both years. Because the use of these elevated grade point averages lead to a predicted performance level for these students that tended to be higher than that actually experienced in the junior and senior class years, several adjustments were made to the forecasting model for predicting major GPA performance.

The first adjustment involved a correction in the Mathematics GPA used by the model for predicting major GPA performance for students who attended military prep school. This adjustment was computed as follows:

	FRESHMAN YEAR MEAN GPA	FRESHMAN AND SOPHOMORE YEARS MEAN GPA
Basic Sciences	2.48	2.46
Social Sciences	2.51	2.56
Engineering Sciences	*	2.59
Humanities	2.67	2.64
Mean GPA - Non-Math	2.55	2.56
Mean GPA Math	2.92	2.85
Correction	- .37	- .29

Thus, the model reduces the Math 4 GPA by .37 in Level B and the Math 4 + 3 GPA by .29 in Level A before using these variables in predicting the major GPA for military prep school attendees.

For non-military prep school attendees, a small adjustment was also made as follows:

	FRESHMAN YEAR MEAN GPA	FRESHMAN AND SOPHOMORE YEARS MEAN GPA
Basic Sciences	2.49	2.54
Social Sciences	2.60	2.60
Engineering Sciences	*	2.43

	FRESHMAN YEAR MEAN GPA	FRESHMAN AND SOPHOMORE YEARS MEAN GPA
Humanities	2.70	2.65
Mean GPA Non-Math	2.60	2.56
Mean GPA Math	2.70	2.66
Correction	- .10	- .10

The asterisk indicates that only 83 out of 301 (28 percent) of the military prep school attendees and 17 out of 65 (26 percent) of non-military prep school attendees had taken any Engineering Sciences courses during the freshman class year. This vas not considered a sufficiently adequate sampling of this variable to include in the correction calculations for the freshman year.

Thus, the model reduces the Math 4 GPA by .10 in Level B and the Math 4 + 3 GPA by .10 in Level A before using these variables in predicting the major GPA for non-military prep school attendees.

Although the Engineering sciences GPA also was out of line for military prep school attendees during the freshman year, a correction for this condition was not made because this variable is not currently used in making Level B predictions. The data for this variable are not available on enough students to consider it for this purpose. No adjustment is required for previous college attendees or non-preps because they are considered as normative students for the purpose of statistical model construction.

Since the equation derivations were all completed before the prep school analysis was completed, the prep school attendees were included in the calculations that produced those equations. Therefore, the equation coefficients assigned to Math 4 + 3 GPA in Level A and to Math 4 GPA in Level B may be somewhat distorted. Other equation coefficients in these models might also have been influenced by the inclusion of these non-normative students in the model. The initial model was, however, mainly determined on the basis of regular students using numerous non-math predictor variables so that the amount of bias due to the prep school predictor-criterion differential should be small.

When the model is updated, all prep school attendees should be deleted. If this should affect sample size for certain majors such that it dips considerably below 50, an alternative approach should be used. That is, the prep school attendee should be retained in the statistical calculations for the given major, but reduce his Math 4 and Math 4 + 3 GPAs by the appropriate corrections described previously.

The Math GPA adjustments described above could disappear entirely or become larger with succeeding classes of graduates. The statistical procedures previously described should be reapplied periodically to monitor this condition and to make further adjustments accordingly.

CACS Validation

As previously noted, the CACS prediction equations were developed using data available on 3234 students who graduated during the time period 1967 through 1971. These equations were subsequently validated on an independent sample of 754 students who graduated in 1972. The purpose of the validation procedure was not only to assess the accuracy of the CACS model, but also to determine the confidence which counselors could place in the model's predictions.

It will be recalled that the model generates predictions for each of 28 majors at three levels of student maturity. The maturity levels are:

Level A - applicable for students that have earned over 45 credit hours.

Level B - applicable for students that have earned 15-45 credit hours.

Level C - applicable for students that have earned less than 15 credit hours.

Validation Procedure

The validation procedure consisted of generating upper division major GPA predictions for the 754 graduating students and comparing such predictions with the students' actual earned GPAs. It was originally planned to use grades accumulated over the last four semesters for computing earned GPAs. However, since final semester grades were not available when the validation test occurred, earned GPAs were, therefore, based on the first three semesters of upper division work.

The comparison of earned versus predicted GPA was made for each major that contained at least one graduate. Only the American Studies major failed to qualify since it had no graduates in 1972. Twenty-seven, rather than 28, prediction GPAs were, therefore, generated for each of the model's maturity levels.

It was apparent from previous analyses that the statistical efficiency of the model's equations would vary from major to major and from maturity level to maturity level. The validation analysis was specifically aimed at assessing this differential prediction capability so that counselors could recognize and use the predictive potential of the model to the best advantage. Additionally, the validation analysis was expected to provide valuable insights regarding avenues for improving the model.

Results of Maturity Level A Validation

Table 7 summarizes the results of comparing the observed (earned) versus predicted GPAs across all 754 students for maturity level A. The average observed GPA was 3.03; whereas, the average predicted GPA was 3.02. This indicates that the distribution of predicted values closely paralleled the distribution of observed values, hence the central tendency calibration of the model was excellent.

Table 7. **Observed and Predicted GPAs for 754 Students (Maturity Level A)**

Student	GPA Observed	GPA Predicted	Individual Difference
1	2.48	2.67	.19
.	.	.	.
.	.	.	.
.	.	.	.
754	3.00	3.32	.32
Overall Average	3.03	3.02	.28

Central tendency calibration is a complex resultant of numerous statistical distributions having an influence on the model. Usually, a model such as CACS will consistently underestimate or overestimate actual performance with the result that the average predicted value will deviate significantly from the average observed value. However, such was not the case for the CACS Level A Model.

The last column of Table 7 shows a second attribute of model validity. The difference between observed and predicted GPAs was computed for each student and then all 754 difference scores were averaged (without regard to plus or minus signs). The average difference score was found to be 0.28 of a GPA, indicating that Model Level A provided a high degree of intrinsic accuracy. Such accuracy it may be noted, is independent of central tendency calibration. That is, it is possible to have high intrinsic accuracy even though central tendency calibration may be quite low, and vice versa.

The distribution of forecasting errors obtained for maturity Level A is summarized in Table 8. The median forecasting error was about 0.23 of a GPA (as contrasted with the mean of 0.28), while the 75th percentile error registered about 0.38 of a GPA. Thus, the bulk of the forecasts fell into the low error category. Only 133 (18 percent) students out of 754 had forecasting errors that might be considered large (*i.e.*, 0.46 of a GPA or higher). (The code numbers for these 133 students have been extracted and preserved for potential follow-up analysis. In such an analysis, all other information available concerning these students would be researched in order to determine possible causations for these errors.)

Comparisons were also made between observed and predicted GPAs for each of 27 majors (Table 9). Columns three and four of Table 9 show the average observed and average predicted GPAs. The overall averages were 2.98 and 2.98, respectively, again indicating high central tendency calibration. Column five presents the differences between observed and predicted GPAs, and the overall average difference is shown as only 0.10 of a GPA. However, the most important measure of intrinsic accuracy consists of averaging difference scores for all students within a major (column 6). The overall average of these scores is shown as 0.27 of a GPA, almost identical to the intrinsic accuracy score obtained in the previous analysis of individual students (Table 7).

Some of the majors had relatively few students, thus, yielding small sample sizes. These are

Table 8. Distribution of Differences Between Observed and Predicted GPAs for 754 Students (Maturity Level A)

(EQUATION A)

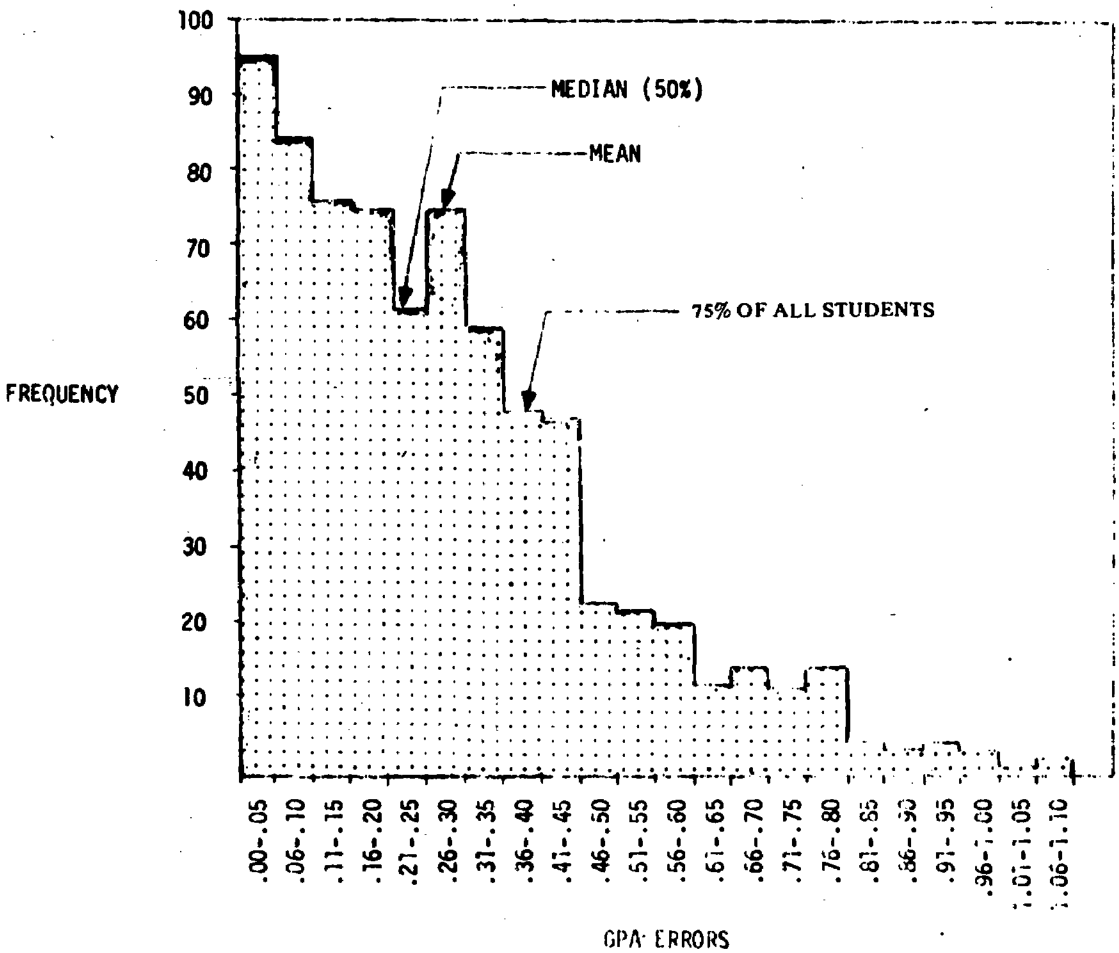


Table 9. Observed and Predicted GPAs for 28 Majors (Maturity Level A)

N[1]	MAJOR		OBSERVED	PREDICTED	DIFFERENCE SCORES	
	NUMBER	TITLE			MEAN	INDIVIDUAL
41	15	Aero. Engrg.	3.19	3.17	.02	.24
Δ	20	Amer. Studies	Δ	Δ	Δ	Δ
23	07	Astronautics	3.30	3.37	.07	.22
17	02	Basic Sciences	2.91	2.59	.32	.42
7	13	Chemistry	3.10	3.11	.01	.24
44	10	Civil Engrg.	3.09	3.10	.01	.34
16	23	Computer Sciences	2.91	2.88	.03	.23
54	17	Economics	3.21	3.09	.12	.33
15	11	Elec. Engrg.	3.23	3.27	.04	.27
32	09	Engrg. Mgt.	2.61	2.75	.14	.26
3	03	Engrg. Sciences	2.77	2.69	.08	.07
4	25	Far East Studies	2.71	2.65	.06	.18
13	22	General Engrg.	2.50	2.72	.22	.38
61	21	General Studies	2.45	2.42	.03	.24
10	19	Geography	2.97	2.90	.07	.43
73	08	History	3.21	3.14	.07	.29
8	01	Humanities	3.00	2.94	.06	.34
57	04	Intl. Affairs	3.06	3.00	.06	.24
2	26	Latin Am. Studies	3.27	3.24	.03	.07
71	24	Life Sciences	3.38	3.28	.10	.25
20	06	Mathematics	3.37	3.38	.01	.29
65	12	Mechanics	3.24	3.30	.06	.32
3	05	Military Science	2.41	2.74	.33	.45
9	18	Political Science	2.49	2.88	.39	.38
23	14	Physics	3.20	3 17	.03	.25
17	16	Psychology	2.98	3.01	.03	.21
5	27	Soviet Studies	2.86	2.82	.04	.21
8	28	W. Europ. Studies	3.04	2.77	.27	.27
751		OVERALL AVERAGE	2.98	2.98	.10	.27
		ADJUSTED OVERALL AVERAGE	3.04	3.03	.08	.29

[1]Keypunch errors resulted in the loss of data for three students, one each from the majors, Aeronautical Engineering, Electrical Engineering, and Life Sciences. Time did not permit correction of these errors. However, these data would not likely provide a perceptible change in the results.

denoted by circles encompassing the major numbers and were arbitrarily selected on the basis of being less than 10. Although such small samples would ordinarily be expected to provide considerable instability, their elimination from the pool of data did not alter the overall averages significantly, as can be seen in the row entitled "Adjusted Overall Average" in Table 9.

Results of Maturity Level B Validation

Maturity Level B forecasting accuracy is summarized in Table 10. The central tendency calibration was offset by only 0.03 of a GPA and the individual differences across all students averaged 0.32 of a GPA. These results correspond rather closely with those obtained for maturity Level A, there being a slight decrease in intrinsic accuracy and central tendency calibration. This outcome is quite satisfactory, considering that Level B sacrifices an entire year of information on a student.

Table 10. **Observed and Predicted GPAs for 754 Students (Maturity Level B)**

Student	GPA Observed	GPA Predicted	Individual Difference
1	2.48	2.71	.23
.	.	.	.
.	.	.	.
.	.	.	.
754	3.00	3.33	.33
Overall Average	3.04	3.01	.32

Table 11 shows the results obtained for the Level B forecast analysis by majors. These results compare favorably with those of Level A, the most important difference being a decrease in intrinsic accuracy of 0.06, on the average; *i.e.*, 0.27 versus 0.33.

Results of Maturity Level C Validation

Table 12 shows the maturity Level C analysis of observed and predicted GPAs. Surprisingly, the central tendency calibration remained high; the offset was only 0.02 of a GPA. However, intrinsic accuracy continued to decrease as expected, the average difference score being 0.38. Since the Level C Model uses no information whatsoever about a student while attending school, such accuracy must be viewed as highly satisfactory.

Table 12. **Observed and Predicted GPAs for 754 Students (Maturity Level C)**

Student	GPA Observed	GPA Predicted	Individual Difference
1	2.48	2.89	.41
.	.	.	.
.	.	.	.
.	.	.	.
754	3.00	3.40	.40
Overall Average	3.04	3.06	.38

Level C forecast analysis by majors is given in Table 13. Again, the central tendency calibration was excellent since the offset was only 0.02 of a GPA. And, as expected, intrinsic accuracy continued to decrease; the overall average deviation was 0.37 of a GPA, 0.10 greater than Level A and 0.04 greater than Level B deviations.

Conclusions

In general, the validation results were quite positive in nature. Central tendency calibrations were consistently high and intrinsic accuracy was excellent, on the average. The former measures lead us to conclude that the CACS prediction model is remarkably free of biases, which would cause it to consistently underestimate or overestimate student performance. With regard to intrinsic prediction errors; *i.e.*, those reflected in individual differences between observed and predicted GPAs, no probabilistic model can hope to eliminate such errors completely. The average individual errors of 0.27, 0.33, and 0.37 for Levels A,B, and C, respectively, appear to be quite tolerable and should not negate the practical value that CACS can provide the school counselors. However, the individual prediction errors in CACS can undoubtedly be reduced through additional analysis. The 133 students who demonstrated the largest deviations from equation forecasts provide a primary research focal point for such analysis.

IV. DISCUSSION AND RECOMMENDATIONS

Overview

The System Development Corporation assembled and provided an initial computer system for the academic counseling.

Table 11. Observed and Predicted GPAs for 28 Majors
(Maturity Level B)

N	MAJOR		OBSERVED	PREDICTED	DIFFERENCE SCORES	
	NUMBER	TITLE			MEAN	Individual
41	15	Aero. Engrg.	3.19	3.12	.07	.30
Δ	20	Amer. Studies	Δ	Δ	Δ	Δ
23	07	Astronautics	3.30	3.43	.13	.29
17	02	Basic Sciences	2.91	2.62	.29	.44
7	13	Chemistry	3.10	3.09	.01	.46
44	10	Civil Engrg.	3.09	3.05	.04	.43
16	23	Computer Sciences	2.91	2.85	.06	.33
54	17	Economics	3.21	3.15	.06	.37
15	11	Elec. Engrg.	3.23	3.30	.07	.32
82	09	Engrg. Mgt.	2.61	2.78	.17	.30
3	03	Engrg. Sciences	2.77	2.52	.25	.27
4	25	Far East Studies	2.71	2.68	.03	.12
13	22	General Engrg.	2.50	2.61	.11	.27
61	21	General Studies	2.45	2.42	.03	.24
10	19	Geography	2.97	2.67	.30	.47
73	08	History	3.21	3.14	.07	.32
8	01	Humanities	3.00	2.98	.02	.39
57	04	Intl. Affairs	3.06	3.01	.05	.27
2	26	Latin Am. Studies	3.27	2.95	.32	.31
71	24	Life Sciences	3.38	3.32	.06	.26
20	06	Mathematics	3.37	3.35	.02	.32
65	12	Mechanics	3.24	3.21	.03	.36
3	05	Military Sciences	2.41	2.72	.31	.46
9	18	Political Science	2.49	2.89	.40	.40
23	14	Physics	3.20	3.19	.01	.31
17	16	Psychology	2.98	3.00	.02	.23
5	27	Soviet Studies	2.86	2.75	.11	.27
8	28	W. Europ. Studies	3.04	2.77	.27	.32
751		OVERALL AVERAGE	2.98	2.95	.12	.33
		ADJUSTED OVERALL AVERAGE	3.04	3.01	.09	.32

Table 13. Observed and Predicted GPAs for 28 Majors
(Maturity Level C)

N[1]	MAJOR		OBSERVED	PREDICTED	DIFFERENCE SCORES	
	NUMBER	TITLE			MEAN	INDIVIDUAL
41	15	Aero. Engrg.	3.19	3.11	.08	.39
Δ	20	Amer. Studies	Δ	Δ	Δ	Δ
23	07	Astronautics	3.30	3.39	.09	.39
17	02	Basic Sciences	2.91	2.58	.33	.49
7	13	Chemistry	3.10	3.17	.07	.47
44	10	Civil Engrg.	3.09	3.07	.02	.46
16	23	Computer Sciences	2.91	2.86	.05	.34
54	17	Economics	3.21	3.03	.18	.43
15	11	Elec. Engrg.	3.23	3.18	.05	.36
82	09	Engrg. Mgt.	2.61	2.98	.37	.43
3	03	Engrg. Sciences	2.77	2.73	.04	.28
4	25	Far East Studies	2.71	3.08	.37	.36
13	22	General Engrg.	2.50	2.61	.11	.34
61	21	General Studies	2.45	2.53	.08	.27
10	19	Geography	2.97	2.80	.17	.42
73	08	History	3.21	3.16	.05	.41
8	01	Humanities	3.00	3.02	.02	.36
57	04	Intl. Affairs	3.06	3.09	.03	.35
2	26	Latin Am. Studies	3.27	3.03	.24	.23
71	24	Life Sciences	3.38	3.28	.10	.36
20	06	Mathematics	3.37	3.37	.00	.28
65	12	Mechanics	3.24	3.27	.03	.42
3	05	Military Sciences	2.41	2.82	.41	.44
9	18	Political Science	2.49	2.86	.37	.44
23	14	Physics	3.20	3.24	.04	.39
17	16	Psychology	2.98	3.06	.08	.30
5	27	Soviet Studies	2.86	2.83	.03	.31
8	28	W. European Studies	3.04	2.90	.14	.33
751		OVERALL AVERAGE	2.98	3.00	.13	.37
		ADJUSTED OVERALL AVERAGE	3.04	3.03	.10	.38

[1]Keypunch errors resulted in the loss of data for three students, one each from the majors, Aeronautical Engineering, Electrical Engineering, and Life Sciences. Time did not permit correction of these errors. However, these data would not likely provide a perceptible change in the results.

This development used the existing data available on students at the school and existing state-of-the-art capability in predictive modeling and interactive computer systems to construct a useable first model that would enhance counseling operations as well as provide capability for self-growth and maturity within the total realm of career counseling.

Objectives

The objectives of the initial Computerized Academic Counseling System (CACS) were to predict the likelihood of success for a student in any of the major fields of study, and to produce a predicted point estimate of grade point average (GPA) for each major and a standard error of estimate value that could be attached to each prediction. This information, along with student supplementary data, was to be made available to an academic counselor in a responsive manner that would enhance his ability to conduct effective counseling.

In general, the objectives of the contracted projects were satisfactorily met, considering the magnitude and quality of data available. An initial mathematical model composed of 84 prediction equations divided into three levels of student maturity was developed and assessed for its academic performance forecasting potential. Also, 31 potential predictor variables available on academic tape files were investigated and converted into weighted equation parameters appropriate to their degree of unique relationship to advanced academic performance.

The amount of performance variation accounted for varied considerably between the three student maturity levels, being 49 percent for the most advanced student Level A, 38 percent for student Level B, and 19 percent for student Level C. However, all of these levels were considered to possess sufficient statistical forecasting efficiency to allow their use by an academic counselor.

The statistical efficiencies of the forecasting models also varied considerably from major to major. Some of this variation was due to the small sample sizes available, for some majors did not allow sufficient stabilization of equations to be achieved. In other cases, the unique nature of the major appeared to be involved; *e.g.*, the Computer Sciences forecasting equations had the poorest statistical forecasting efficiency in all levels.

Design Rationale

The five basic design criteria of long life expectancy, ease of expansion, ease of updating, simplicity of use, and timely response were met quite well, considering the relative complexity of the various components that went into development of the system and the accelerated production schedule that had to be maintained in order to achieve the installation date. The major design limitations that can be improved upon are as follows:

1. Increasing the sample sizes used in equation derivation, especially for those majors that tend to have fewer graduates.

2. Adding additional pertinent variables for each student that reflect such factors as motivation, vocational interest, personality, and specific area aptitude/achievement indicators.

3. Augmenting the academic success forecasting model with a parallel model based on group membership probability.

4. Creating additional models that will tie the academic performance at the school with industry manning and replacement factors.

5. Modifying the hardware/software aspects of the system and its supporting elements to efficiently route a greater abundance of data to counselors as well as modeling analysts.

System Expansion

There are three principal areas of potential system expansion. One is in the direction of enlarging the scope of academic counseling through use of additional mathematical models, hardware, and software to encompass the larger sphere of career counseling within the manning requirements of industry.

A second area concerns the improvement of the efficiency of forecasting performance at the school by developing additional models while existing models are improved through augmented data collection and analysis.

A third area involves the computer efficiency of the hardware and software systems both within CACS and the environment in which it can best be expected to operate. New display devices, capabilities and interactive computer system concepts which are constantly being developed should be reflected in the CACS system expansion.

Recommendations for Future Research

In the scientific literature on personnel classification and counseling; *e.g.*, (Bennett, Seashore, & Wesman, 1952; Cattell, 1949; DuMas, 1949; Dunn, 1955; Rulon, 1967; Stewart, 1947; Strong 1931; Tatsuoka, 1956), supplementary mathematical models are discussed that serve to complement the typical success forecasting model. The most prominent of these has been referred to as the group membership modeling approach, designed to determine the extent to which a person possesses those particular characteristics that are associated with typical members of a given occupational group. The forerunner of this approach was the excellent work performed by Strong (1931) with respect to vocational interests.

Group membership modeling theory argues that it is not meaningful to estimate the likelihood of success in a given occupational area unless the person can legitimately be considered a competitive member of that occupation; *i.e.*, to have a profile of abilities, interests and personality characteristics that fit well with the current members of the occupation.

Frequently, a specific profile characteristic may be very pertinent to a given field of endeavor yet show no statistical relationship to performance in that field. In her study of Brown University majors, Dunn (1955) found that her equations for predicting grade point averages of chemistry majors gave no weight to mathematics ability while her equations for history majors gave no weight to verbal ability. She concluded that individuals within these majors had self-selected themselves and; consequently, became highly homogeneous on each characteristic. As a result, these two types of abilities did not differentiate students within these majors. Apparently the variance phenomenon, called restriction of range of talent, operated on the statistical correlation between mathematical/verbal factors and chemistry/history grade performance, in the Dunn studies, in such a manner as to attenuate the impacts in the equation model. Dunn also found that mathematics ability picked up a weight for history majors. This demonstrates how secondary predictors can assume more weight in an equation than primary factors, due to the fact that the correlational influence of the primary predictors has been homogenized out by selective factors, while the influence of the secondary predictors has not.

For advanced modeling purposes applicable to this Computerized Academic Counseling System, as well as, the general state-of-the-art in computerized occupational guidance, the following considerations are proposed.

1. Augment the 31 present predictor variables in CACS with suitable vocational interest and personality measures for each student. The measures chosen should have the quality associated with recognized instruments like the Strong Vocational Interest Blank and the Minnesota Multi-phasic Personality Inventory.
2. Conduct a thorough factor analysis of all predictors so as to reduce the effective statistical dimensionality of the independent variable domain.
3. Define major occupational groups into which all school attendees can be grouped. These groups should be sensitive to the existing vocations in industry and be tied to given majors or clusters of majors. The groups may include past graduates as well as undergraduates who are essentially committed to a given field of endeavor, such as engineering.
4. Combine all students into one large sample for further analysis. Each student will have a series of dummy dependent variables showing zeros except for the group to which he belongs.
5. Conduct systematic multiple regression/discriminant analyses using the group membership dummy variables with respect to the major occupational groups as dependent variables and the factors isolated in item 2 as independent variables. Caution: Before an independent variable is considered for inclusion in this model, it must contain data values for all students or at least for most all students; otherwise, non-Gramian statistical effects will result giving misleading group membership matching equations.
6. Obtain group membership matching equations with the weights assigned to each factor so as to maximize the homogenity of each group as well as its distinction from all other groups.
7. Convert weights from factors back into terms of the original measures used as independent variables. This will provide a more visible method by which the specific effects of each variable in the basic data array can be assessed.
8. Create an algorithm for converting equation results to a probability statement that a given student possesses the attributes befitting each major occupational group in the complementary model.

This algorithm can be obtained in the manner described in the Success Probability Section of the CACS Mathematical Model Description in the appendix.

REFERENCES

Bennett, G.K., Seashore, H.G., & Wesman, A.G. Aptitude testing: Does it 'Prove out' in counseling practice? *Occupations*, 1952, **30**, 584-598.

Burket, G.R. A study of reduced rank models for multiple prediction. *Psychometric Monographs*, 1964, No. 12.

Cattell, R.B. RP and other coefficients of pattern similarity. *Psychometrika*, 1949, **14**, 279-298.

DuMas, F.M. The coefficient of profile similarity. *Journal of Clinical Psychology*, 1949, **5**, 123-131.

Dunn, F.E. Guiding college students in the selection of a field of concentration. Unpublished doctoral dissertation, Harvard, 1955.

Herzberg, P.A. The parameters of cross-validation. *Psychometrika Monograph Supplement*, 1969, No. 16.

Horst, P. *The prediction of personal adjustment.* New York: Social Science Research Council, 1941.

Leiman, J.M. The calculation of regression weights from common factor loadings. Unpublished doctoral dissertation, University of Washington, 1951.

Rulon, P.J. *Multivariate statistics for personnel classification.* New York: Wiley, 1967.

Stewart, N. AGCT scores of Army personnel grouped by occupation. *Occupations*, 1947, 24, [illegible].

Strong, E.K. *Vocational interest blank for men.* Stanford University Press, 1931.

Tatsuoka, M.M. Joint Probability of membership in a group and success therein. Unpublished doctoral dissertation, Harvard, 1956.

APPENDIX: MATHEMATICAL MODEL DESCRIPTION AND MAINTENANCE

Introduction

This report constitues one of a series of documents describing the design, development and evaluation of a Computerized Academic Counseling System (CACS). The present document describes the technical development of a mathematical model which represents the core of CACS. The purpose of the model is to predict, during the first two years of academic work, eventual grade point average (GPA) and likelihood of achieveing at least a 2.0 GPA for each student in any of 28 major fields of study. The model is primarily composed of a set of multiple linear regression equations which utilize input data routinely collected.

A discussion of the mathematical model, as it relates to other components of the CACS system, is presented in the official R & D contract file.

The Mathematical Model for Grade Point Estimation

The mathematical model for predicting academic success is designed to generate grade point estimates for a student at three different levels, depending on how far he has advanced in his school work. Level A is designed to forecast success for relatively advanced students, Level B for moderately advanced students, and Level C for relatively new students. All levels are focused on students in their first two years with the dependent variable being the major GPA likely to be achieved during the last two years. Separate mathematical equations are provided within each level for 28 majors that were active during the time period 1967 through 1971. The mathematical form used to generate the point estimate within each level is as follows: $BX + C = Y$

where $B = \begin{matrix} b_{1,1} & b_{1,2} & \cdots & b_{1,31} \\ \vdots \\ b_{28,1} & b_{28,2} & \cdots & b_{28,31} \end{matrix}$ is a 31 by 28 matrix of linear regression coefficients. This matrix will contain null coefficients for those predictor variables that are not currently used in making a point estimate for a given major. See the System Specification (official R & D contract file) for the values of the coefficients currently stored in the program.

$X = \begin{matrix} x\,1 \\ x\,2 \\ \vdots \\ x\,31 \end{matrix}$ is a vector of 31 numerical predictor values for the student being counselled. Similar vectors are compiled from the CAIDS-PRC data base for active students and are made available to the model on call. All 31 predictor values are available for counselor use, although only a portion of these are actually used in making the grade point estimates.

$C = \begin{matrix} c\,1 \\ c\,2 \\ \vdots \\ c\,28 \end{matrix}$ is a vector of 28 regression constants uniquely applicable to each major. These constants are part of the regression equations (the y-intercepts) computed during statistical analysis.

$Y = \begin{matrix} y\,1 \\ y\,2 \\ \vdots \\ y\,28 \end{matrix}$ is a resultant vector of 28 grade point estimates, one for each major for the student being processed. If only one major of interest is specified by the counselor, the 27 remaining estimates are bypassed by the program.

The B matrix filed in the CACS Prediction Module contains provisions for 31 prediction equation coefficients for each major. This matrix table is flexible in that any one of the 31 variables forwarded from the Data Retrieval Module can be weighted. Thus, there is considerable room for equation expansion without the necessity to overhaul the Prediction Module. In the current mathematical model, however, only 16 of the 31 independent variables are actually weighted in the major GPA prediction equations. The remaining 15 variables were nonweighted because of the failure to survive the variable selection criteria described in Steps 10 through 16 of the Updating Procedures in the discussion paragraph, page 22 (*e.g.*, lack of correlation with major GPA, redundancy with a statistically superior variable).

After an updated statistical analysis is conducted, regression coefficients for existing predictors can be easily changed without redesigning the program.

Standard Error of Estimate

Along with the major GPA performance point estimates provided by the elements of the Y-vector, the CACS model also provides standard errors of estimate to be used by the counselor in assessing the relative confidence that may be placed in the forecasts. The standard error values are computed at the same time as the regression equations and are stored in the program as an E-vector containing 28 elements to correspond with the Y-vector. Each E-element is defined by the following equation:

$$E_i = \left[\frac{\Sigma(Yoi\text{-}Yci)^2}{N\text{-}M\text{-}1} \right]^{1/2}$$

where E_i is the standard error of estimate corresponding with the ith of the Y-vector.

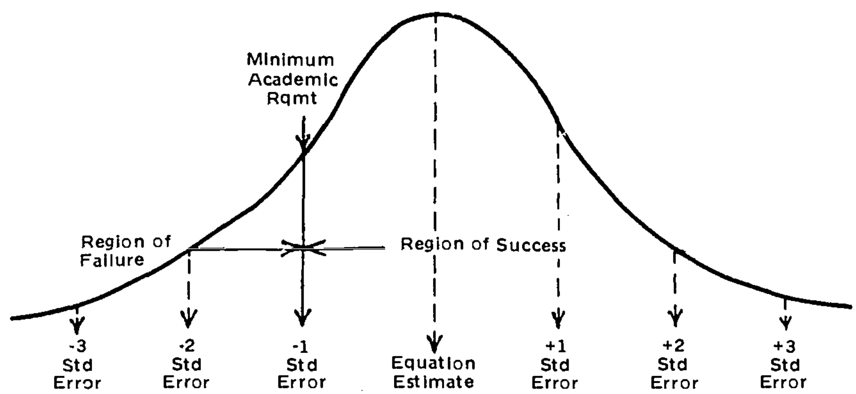
> Yoi is the observed major GPA compiled for each student graduate in the data base used for equation derivation.
>
> Yci is the computed value of the major GPA when applying the equation to each graduate in the data base from which it was derived.
>
> N is the number of student graduates in the data base used for equation derivation for the specific major.
>
> M is the number of predictors used in the equation to predict Yi.

It is assumed that the computed standard error will remain comparable for new students. This assumption, however, is contingent on the size of the data base used, the number of predictors applied, and the consistency of relationships (correlations) among the predictors and the criterion; *i.e.*, major GPA. A special study is being conducted to test the consistency of the standard error on new students. The results of this study may lead to an operational correction in the computed standard error in order to make it more realistic in actual counseling.

Success Probability

Along with each point estimate and accompanying standard error of estimate, the CACS model involves several mathematical transformations to arrive at a determination of success probability on a scale from .01 to .99.

The approach to the problem of estimating the likelihood of success for a student in a given major was defined on the basis of the standard statistical forecasting theory, which is summarized as follows. The estimated major GPA and the standard error of estimate are taken as the parameters (mean and standard deviation) defining the most likely distribution of prediction errors for each forecast. The minimum academic requirement for acceptable major GPA is set into the context of this distribution, and standardized in terms of the normal unit curve. This can be illustrated as follows:



The portion of the standard normal curve to the right of the minimum academic requirement represents the probability that the actual major GPA will fall into the acceptable region. The size of this region relative to the total curve becomes the probability of success. In the illustrated example, the minimum academic requirement falls at one standard error of estimate below the estimated major GPA, therefore, the region of success represents 84 percent of the total curve. The region of failure, i.e., the likelihood that the actual GPA will be below minimum requirements represents 16 percent of the total curve. Thus, the probability of success is set at .84.

The computational algorithm for determining this probability is as follows:

$$\text{trial probability} = 1\text{-}½\left(1+C_1\left|Z\right|+C_2\left|Z\right|^2+C_3\left|Z\right|^3+C_4\left|Z\right|^4\right)^{-4}$$

where

$$C_1 = .196854$$
$$C_2 = .115194$$
$$C_3 = .000344$$
$$C_4 = .019527$$

If $o > Z \geq -2.17$, probability = 1.00 minus trial probability.

If $o \leq Z \leq 2.17$, probability = trial probability.

If $Z > 2.17$, probability = .99 and trial probability is not computed.

If $Z < -2.17$, probability = .01 and trial probability is not computed.

A standard ratio defining the deviation of the point estimate from minimum accepatable performance is given by:

$$Z_i = \frac{Y_i - 2.0}{E_i}$$ (Limits of 0-4 are set for Y_i in this calculation.)

where Y_i = the point estimate of major GPA obtained as an element in the Y-vector previously described.

2.0 = a constant describing the minimum acceptable major GPA at the school.

E_i = the standard error of estimate corresponding to the point estimate of the ith element in Y.

The success probability provides an estimate of the likelihood that the student being counselled will be able to achieve at least a minimum grade point average of 2.0 in the specified major.

Data Base

The data base for the update is derived from the CAIDS-PRC data files for past graduates of the school starting with the class of 1967. The data set for each graduate consists of:

1.	Student Control Number	CAIDS Item MXC2, tape positions 863-868
2.	Major Code	CAIDS Item MXM1, tape positions 427-428
3.	Potential Predictor Variables	As described in the potential predictor variables paragraph
4.	Major GPA (Variable 32)	As described in the major GPA (variable 32) paragraph.

Potential Predictor Variables

Variable Number	Source	Statistical Limts	Description
1	PRC	1.00-4.00	Grade point average (GPA in Mathematics during fourth and third class years. Calculate the GPA for all Mathematics courses during fourth and third years. This consists of all courses with a Department Code Letter of "P" in column 51 on "G" and "A" records.
2	PRC	1.00-4.00	GPA in other Basic Sciences during fourth and third class years. These courses are those with a Department Code letter of "E", "T", or "S".
3	PRC	1.00-4.00	GPA in Engineering Sciences during fourth and third class years. These courses are identified with Department Code Letters "B", "C", "F", "H", and "R".
4	PRC	1.00-4.00	GPA in Humanities during fourth and third class years. These courses are identified with Department Code Letters "J", "K", and "M".
5	PRC	1.00-4.00	GPA in Social Sciences during fourth and third class years. These courses are identified with Department Code Letters "G", "L", "U", "N", and "D".
6			GPA in Mathematics during fourth class year. (Otherwise same as Variable 1.)
7			GPA in other Basic Sciences during fourth class year. (Otherwise same as Variable 2.)

Variable Number	Source	Statistical Limits	Description
8			GPA in Engineering Sciences during fourth class year. (Otherwise same as Variable 3.)
9			GPA in Humanities during fourth class year. (Otherwise same as Variable 4.)
10			GPA in Social Sciences during fourth class year. (Otherwise same as Variable 5.)
11	CAIDS	0-1	Falcon or Skelly Scholarship. Item MXFS (Pos. 18). If item contains an "F" or "S", this variable will have a value of "1"; "0" for all other students.
12	CAIDS	0-2	Turnback Indicator. If item MXD3 (Pos. 72, 73) contains a year, this variable will have a value of "2". If item MXD? (Pos. 70,71) contains a year and item MXD3 is blank, this variable will have a value of "1". If both above items are blank, this variable will have a value of "0".
13	CAIDS	20-25	Age at Graduation. Pos. 442, 443 of item MXT2 minus Pos. 93, 94 of item MXDB.
14	CAIDS	400-800	Prior Academic Achievement. Item MXPR (Pos. 797-799).
15	CAIDS	400-800	Verbal Aptitude. Item MXVA (Pos. 800-802).
16	CAIDS	400-800	English Composition. Item MXEN (Pos. 803-805).
17	CAIDS	900-1600	Composite English Score. Item MXEC (Pos. 806-809).
18	CAIDS	500-800	Math Aptitude. Item MXMA (Pos. 810-812).
19	CAIDS	1-2	Intermediate or Advanced Math Code. Item MXIA (Pos. 813). If coded "A", convert to "2"; if "I", convert to "1".

Variable Number	Source	Statistical Limits	Description
20	CAIDS	500-800	Math Achievement. Item MXMV (Pos. 814-816).
21	CAIDS	1000-1600	Composite Math Score. Item MXMC (Pos. 817-820).
22	CAIDS	2600-4000	Academic Composite. Item MXAC (Pos. 821-824).
23	CAIDS	400-800	PAE Score. Item MXPA (Pos. 825-827).
24	CAIDS	300-800	Activities - Athletic. Item MXAA (Pos. 828-830).
25	CAIDS	300-800	Activities - Nonathletic. Item MXAN (Pos. 831-833).
26	CAIDS	1200-2400	Leadership Composite. Item MXLD (Pos. 834-837).
27	CAIDS	500-800	Weighted Composite. Item MXWC (Pos. 838-841). Delete the least significant digit shown in the student Master Tape. Then all variables obtained from this tape will be scaled as integers.
28	CAIDS	1-4	Medical Qualification Code. Item MXMD (Pos. 842). If this column contains a letter, convert to a number as follows: A = 1 S = 1 B = 2 T = 2 C = 3 U = 3 D = 4 V = 4 E = 5 W = 5 F = 6 X = 6 All other letters = 7
29	CAIDS	0-1	Military Prep School Attended Code. Item MXPP (Pos. 843). If item contains an "A", convert to "1"; all others convert to "0".
30	CAIDS	0-1	Other Prep School Attended Code. Item MXPP (Pos. 843). If item contains a "P", convert to "1"; all others convert to "0".

Variable Number	Source	Statistical Limits	Description
31	CAIDS	0-1	College Attended Code. Item MXCO (Pos. 844). If item contains a "C", convert to "1"; all others convert to "0".

Major GPA (Variable 32)

The procedure is designed to obtain major GPA during the second and first class years. The Academic curriculum handbook for 1971 through 1972 was used as a guideline in developing this procedure.

The student Master File is merged with the PRC File to create a card image data file for each year starting with 1967. The merged file is arranged in ascending order by student identification number for both graduates and nongraduates. Then, the 31 potential predictor variables and major GPA are calculated for all students except those that have no: (*a*) CAIDS data, (*b*) PRC data, (*c*) graduation indication (Item MXT2, Pos. 437-443), and (*d*) major indication (Item MXM1, Pos. 427-428).

In calculating the major GPA, the "Y" card is used to determine the second and first class years. The "G" and "A" cards following the "Y" card provide the courses, grades, and hours credited the student during the second and first class years.

The information obtained from the "G" and "A" card is as follows:

1.	Col. 51	Department Code Letter, which identifies the course
2.	Col. 53	Letter Grade
3.	Col. 55-58	Credit Hours (3 decimal places)

The Department Code Letter found in Col. 51 is compared to the course code determined for that major. Table 14 contains the 28 majors, the two-digit school Code, and the course code within each major. If the Department Code found on the "G" and "A" card is a Department Code defined for that major in Table 14, then the grade data for the course is included in the calculation for that major GPA, otherwise it is ignored.

The computational formula is:

$$Y = \Sigma HG / \Sigma H$$

where Y = Major GPA (Statistical Limits: 1.00-4.00)

H = credit hours for the course

G = converted letter grade (A=4, B=3, C=2, D=1, F=0; all other letter grades are ignored)

Updating Procedures

The following steps describe the mathematical model updating procedures.

Step	Procedure
1	The data set for each graduate is compiled from the CAIDS-PRC files for class years 1967 through 19xx.

Table 14. **Distribution of Course Codes by Majors.**

Major	2-DIGIT SCHOOL CODE	Chemistry	Life Sciences	Mathematics	Physics	Aeronautics	Astronautics & Computer Sc	Civil Engrg	Elec Engrg	Engrg Mech	English, Philosophy, Fine Arts	Foreign Language	History	Economics and Management	Geog	Law	Pol Sc	Psych, Leadership	Sciences**	Area Studies**	Navagation	Airmanship*
Aeronautical Engrg	15			P	S	B	C	F	H	R									V**			
American Studies	20										J	K	M	G	L	N	U	D		V**		
Astronautics	07			P		B	C	F	H	R									V**			
Basic Sciences	02	E	T	F	S														V**			
Chemistry	13	E																				
Civil Engrg	10							F		R												
Computer Sciences	23			P			C															
Economics	17													G	L	N	U	D		V**		
Elec Engrg	11								H													
Engrg Mgt	09													G		N		D				
Engrg Mech	12			P						R												
Engrg Sciences	03	E	T	P	S	B	C	F	H	R									V**			
Far Eastern Studies	25										J	K	M	G	L	N	U	D		V**		
General Engrg	22					B	C	F	H	R									V**			
General Studies	21	E	T	P	S	B	C	F	H	R	J	K	M	G	L	N	U	D	V**	V**	W	X-400
Geography	19														L							
History	08												M									
Humanities	01										J	K	M							V**		
Internatl Affairs	04										J	K	M	G	L	N	U	D		V**		
Latin Amer Studies	26										J	K	M	G	L	N	U	D		V**		
Life Sciences	24		T																			
Mathematics	06			P																		
Military Arts & Sc	05	E	T	P	S	B	C	F	H	R	J	K	M	G	L	N	U	D	V**	V**	W	X-400
Political Science	18																U					
Physics	14			P	S																	
Psychology	16																	D				
Soviet Studies	27										J	K	M	G	L	N	U	D		V**		
W European Studies	28										J	K	M	G	L	N	U	D		V**		

*If Col. 51 contains an "X", check that Cols. 20-22 contain a 400; ignore all other "X" courses.

**If Col. 51 contains a "V", check Cols. 12-19. Area Studies courses will contain AREA STU and Science courses will contain SCI in these columns.

Step	Procedure
2	All the above data sets are sorted and counted by graduation years within each major.
3	The mean major GPA for each major for each year is computed.
4	Any sharp change in mean GPA, that can be attributed to some factor other than sampling fluctuation, is studied for possible implications that could lead to the deletion of certain years from the subsequent calculations for a major. Drastic changes in grading standards or requirements for a major are among the potential causes for this type of action. Another could be the deactivation of a major, which has had graduates in the past. Still another could be insufficient data for a major to permit confidence in the equations derived. There is no absolute rule for dictating what the minimum number of students should be before an equation solution for a major is attempted. SDC experience recommends a sample size of approximately 50 or more students as a reasonable working limit. Exceptions to this rule may have to be made to keep all majors in the model.
5	After the student data sets are screened for inclusion/exclusion of graduating years and majors, they are grouped by majors for further testing concerning prep school attendance. Special SDC studies have revealed that the actual values for certain predictor variables used by CACS are inflated by prep school attendance. The variables isolated so far are GPA-Mathematics fourth class year, and GPA-Mathematics fourth and third class years. The grade bias in these variables is revealed when computing the fourth class year and the fourth plus third class years mean GPA for graduates who attended prep school and those who did not in the following subject matter fields: Basic Sciences (other than Math), Social Sciences, Engineering Sciences, Humanities, and Mathematics. The mean GPA in Mathematics for prep school attendees has been shown to exceed the composite mean GPA relative to the other four subject matter areas. A correction in the CACS program and an adjustment in the data base used for model construction is recommended in order to reduce the effect of this bias on the CACS predictions and future equations to be computed. The preferred procedure is to eliminate all prep school data sets from inclusion in future regression computations because they serve to distort the true effect of the mathematics variables and perhaps other variables on major GPA performance. If the deletion of prep school attendees works a hardship on the sample size for a major that does not produce many graduates, an alternative procedure may be invoked. Determine a correction in the mathematics grade scores used by CACS for predicting major GPA performance for students who have attended military prep school. This correction is computed as follows:

	4th Class Year Mean GPA	4th & 3rd Class Years Mean GPA
Basic Sciences	2.48	2.46
Social Sciences	2.51	2.56
Engineering Sciences	*	2.59
Humanities	2.67	2.64
Mean GPA Non-Math	2.55	2.56
Mean GPA Math	2.92	2.85
Correction	– .37	– .29

Step	Procedure

Thus, CACS would reduce the Math 4 GPA by .37 and the Math 4 + 3 GPA by .29 before using these variables in predicting the major GPA for a student who attended military prep school.

For a student who attended a non-military prep school, a smaller adjustment would be made as follows:

	4th Class Year Mean GPA	4th & 3rd Class Years Mean GPA
Basic Sciences	2.49	2.54
Social Sciences	2.60	2.60
Engineering Sciences	*	2.43
Humanities	2.70	2.65
Mean GPA Non-Math	2.60	2.56
Mean GPA Math	2.70	2.66
Correction	– .10	– .10

The asterisk indicates that this major is not used in 4th class year correction due to lack of data for this variable on most students.

Thus, CACS would reduce the Math 4 GPA by .10 and the Math 4 + 3 GPA by .10 before using these variables in predicting the major GPA for a student who attended a non-military prep school.

There would be no adjustment for previous college attendees or non-preps as they would be considered as normative students for the purpose of the model.

As a result of including the prep school attendees in the initial CACS model derivations, the statistical computations used to derive the coefficients assigned to Math 4 + 3 GPA in Level A and to Math 4 GPA in Level B were somewhat contaminated. This contamination may have extended in lesser degree to other coefficients that were influenced by the inclusion of what was regarded as non-normative types of student graduates in the derivations. Since the initial model was statistically determined on the basis of a preponderance of acceptable normative graduates and numerous non-math predictor variables, it was not considered cost-effective to perform a recalculation of the initial model.

However, when the model is updated with new graduates, any graduate who attended any type of prep school should be deleted from the statistical computations used to derive the equations. If this procedure works a hardship on the sample size for certain majors such that it dips considerably below 50, an alternative approach may be used. Here one would keep the prep school attendee in the statistical calculations for the given major but would reduce his Math 4 and Math 4 + 3 GPAs by the appropriate corrections described earlier.

It is possible that the predictor variable adjustments described above could disappear or become larger with succeeding classes of graduates. Therefore, the statistical correction procedures should always be applied during CACS mathematical model maintenance.

For the next step, it will be assumed that all inappropriate graduated student data sets will either have been eliminated from the data base or suitably corrected as described above.

Step	Procedure

6 Obtain data listings by majors for each variable in the data set. These listings are screened to assure that majors are not mixed and that reasonable data values are available for each variable. The statistical limits shown in the previous section can be used as guidelines for screening data values.

7 Using a conventional product moment correlation analysis program, compute for each major, a 32 by 32, variable pair by variable pair, symmetric correlation matrix (R) where each element rij is defined as follows:

$$rij = \frac{N_{ij}\ \Sigma X_i X_j - \Sigma X_i\, \Sigma X_j}{\left[N_{ij}\ \Sigma X_i^2 - \left(\Sigma X_i\right)^2\right]^{1/2} \left[N_{ij}\ \Sigma X_j^2 - \left(\Sigma X_j\right)^2\right]^{1/2}}$$

where i is the row subscript ($1 \leq i \leq 32$)

j is the symmetric column subscript ($1 \leq j \leq 32$)

N_{ij} is the number of paired values for the element being computed

It is desirable in the above calculation that the subscript 32 be assigned to the dependent variable for the model, which is major GPA. This will make it easier, along with the symmetric format, to observe the respective correlations of each independent variable with major GPA and will expedite the screening of independent variables as potential predictors. When the N_{ij} for computing an element in the R matrix falls considerably below 50, that element and the data variable responsible for this condition should be flagged for caution and possible exclusion from final equation derivation.

8 Using the 28 correlation matrices computed in Step 7 (there is a different one for each major) and a standard linear regression analysis program, compute for each independent (predictor) variable within each major and level:

a. The standardized partial regression coefficient (beta coefficient) with respect to the dependent variable, major GPA

b. the correlation coefficient with the dependent variable

c. the product of the beta and correlation coefficients

d. the correlation coefficient, with every other nominated independent variable

Also for each equation, compute the mean, standard deviation, and unstandardized B-coefficient for each independent variable; the regression constant (Y-intercept), the coefficient of determination, the multiple correlation coefficient, and the standard error of estimate; and the mean, standard deviation and sample size for the dependent variable (major GPA).

In the first pass through the regression analysis and trial equations that will result, nominate as independent variables only those predictors currently having non-null coefficients in the B-matrix for each level. If all 28 majors, that were active during the 1967 through 1971 period,

Step	Procedure
	are preserved in the analysis, there will be a total of 3 levels times 28 equations within each level or a total of 84 regression analyses in the first pass of the updating process.
9	Examine the outputs of each regression analysis to determine whether any current predictor variable should be eliminated from subsequent passes through the analysis. The criteria for eliminating variables will be described in Steps 10 through 14.
10	First make a correlation sign check to determine if the direction of the correlation of each nominated independent variable is in the proper direction. All correlation signs with respect to major GPA should be positive except for the following variables, which have a logical negative relationship with the dependent variable.

a. estimated age at graduation

b. turnback indicator

c. medical qualification code

d. Falcon/Skelly scholarship fund

e. military prep school attendance indicator

f. other prep school attendance indicator

Note - The variables listed in b through f were not used as predictors in the initial model.

11 Next, make a correlation magnitude test. To exceed chance occurrence probability, the magnitude should be greater than 2 divided by the square root of N_{ij}, where N_{ij} is the number of paired data values used to compute the correlation coefficient in question. The correlation magnitude criterion may be raised still higher at the discretion of the analyst in the event that there are numerous nominated independent variables that pass the minimum magnitude test.

12 For all independent variables that pass the correlation sign and magnitude test, apply the standardized partial regression coefficient (beta coefficient) sign check. The sign of the beta coefficient should correspond with the sign of the correlation coefficient. If it does not, there are probably overlapping predictor variables in the regression analysis. One or more variables should be eliminated to remove the overlap. The choice of the variable(s) to be removed in this case depends on extra-statistical considerations.

13 After the beta sign check, examine the standardized partial regression coefficients for magnitude. Ideally, the magnitude of the beta coefficient should approximate the magnitude of the correlation coefficient, if all predictors were truly independent of each other. However, due to partially overlapping variables, the magnitude of a beta coefficient can fall considerably below the correlation coefficient even at times, assuming an illogical sign, when the overlap is high relative to the correlation of the competing independent variables with the dependent variable.

When a relatively large number of independent variables is nominated and the N_{ij} on which the correlations for these variables is based falls considerably below 50, the beta coefficients may assume distorted magnitudes that are considerably higher than those of the correlation coefficients. This condition often disappears as the number of independent variables is reduced

Step	Procedure

or the sample size is increased. No absolute rule exists for matching the appropriate number of independent variables to sample size. However, SDC experience recommends that no more than one predictor be employed for each ten student graduate data sets used to derive an equation.

For the beta magnitude test, a useful approximation is the one described in Step 10 for the correlation magnitude; namely, 2 divided by the square root of N_{ij}, where N_{ij} is the number of paired values used to compute the correlation coefficient from which the beta coefficient is derived. If the independent variable has a beta coefficient that is less than this magnitude, it probably will have little impact on the dependent variable to be worth including in the equation. Sometimes; however, variables like this are left in because they may have considerable face validity for an equation. If the variable has a sizeable correlation coefficient and a moderately low beta coefficient, it is desirable to keep it in the equation, since it will have a combined impact that is not trivial as will be seen in the next step.

14 For variables that survive the correlation/beta coefficient sign and magnitude tests, apply the correlation x beta coefficient product test. As implied in the previous steps, the sign of the product should always be positive, which indicates a consistency of signs for both the correlation coefficient and the beta coefficient. No absolute rule exists concerning the magnitude of the product to be desired. However, SDC experience recommends a minimum value of .01 in the product of the two coefficients. Such a value corresponds to a 1 percent impact of the predictor variable in the statistical determination of the dependent variable, namely, major GPA performance.

15 After the model equations are purged of any undesirable predictors in Steps 10 through 14, the CACS model analyst may introduce and test previously unused predictor variables into the competition for space in the regression model. Ideally, variables should be introduced into an existing equation one at a time because the resulting effects on the regression coefficients are easier to observe. The regression coefficient for a predictor variable is mathematically sensitive to what other predictor variables are included in the trial equation solution. If many new variables are introduced simultaneously, it becomes difficult to isolate the particular effects produced by specific variables.

16 For economic as well as scientific parsimony considerations, new variables should be carefully screened before attempting to introduce them into a regression equation. There are four basic criteria used for such screening. These are:

a. *Appropriateness of Variable*. Is the variable an appropriate predictor for the level of the mode' being analyzed? For example, Mathematics GPA accrued during the fourth class year is an appropriate predictor for a Level B equation, but not for a Level A or Level C equation. For Level A, Mathematics GPA accrued during the fourth and third class years is undoubtedly a superior predictor. For Level C, neither predictor is suitable, although they can be compiled for all graduates, because they will not be available for a new student during his first semester when the Level C equations must be applied.

b. *Data Quality*. Is data available and usuable as a predictor for majority of the students? Is the data reliable and based on a proven measurement device and procedure? Is the data quantitative rather than qualitative in nature? Is there confidence in what the data purport to measure; *i.e.*, would different analysts ascribe similar significance to high, medium, and low data values?

c. *Correlation with Performance*. Does the variable correlate beyond chance expectation with major GPA? Does it correlate higher than those predictors already in the equation? The correlation matrix computed in Step 7 can be consulted for this purpose.

d. *Independence as to Current Predictors.* Does the variable correlate substantially less with the current predictors than with major GPA? The correlation matrix computed in Step 7 can be consulted for this purpose. If the new variable is highly correlated with existing predictors it will, when introduced into a trial solution, contribute little or nothing to the composite accuracy of the equation and may in fact disturb an existing solution so that illogical signs and distorted magnitudes in beta coefficients will appear.

Because a substantial number of the variables in the CACS data set are composites of other variables, caution should be exercised before introducing a composite when one or more of its components or a similar variable is already in the existing equation model. Criteria c and d should be carefully applied, and if it is still desirable to attempt to introduce the composite into the model, it should be done one variable at a time while observing what happens to the regression coefficients already in the equation. This procedure would be recommended, for example, if the analyst desired to introduce a variable like the Composite English Score into an equation already containing Verbal Aptitude.

Discussion

The foregoing description of CACS mathematical model maintenance is based on standard regression analysis principles that have been proven for many years. The application of these principles should lead to satisfactory service from the model. However, there are other approaches to model construction that may be considered in this context. Some of these may appear to be fascinating potential alternatives to the standard regression approach, but caution must be exercised before an alternative approach is substituted for the one currently installed. The comparison of assets and liabilities of competing approaches should be carefully investigated by thorough and conclusive scientific testing based on such criteria as relative cost, accuracy, and efficiency.

For example, a model could be constructed that forced a prediction weight for all or most of the 31 predictor variables in the student data set. This has certain appealing features associated with the redundancy of the data provided by the CAIDS-PRC files. Presumably, the model would use all of the data that is available to it. The equation weights for such a model could be derived by a combination of factor analysis and regression analysis procedures similar to those discussed (Burket, 1964; Herzberg, 1969; Horst, 1941; Leiman, 1951).

Conceivably, a model like this might be somewhat more accurate than the standard regression model, especially when the sample size for a major is relatively small. Whether the net accuracy accrued across all majors and levels would be justified by the additional data analysis upkeep incurred is unknown at present. Only a special and thorough investigation based on rigorous, comparative analysis will be able to answer this question or related questions concerning other modeling approaches that might appear to have potential.

Unclassified

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY *(Corporate author)* System Development Corporation 2500 Colorado Avenue Santa Monica, California 90406	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP N/A

3. REPORT TITLE

AUTOMATIC DATA PROCESSING SYSTEM AND PROCEDURES
COMPUTERIZED ACADEMIC COUNSELING SYSTEM

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*

Final Report

5. AUTHOR(S) *(First name, middle initial, last name)*

Henry J. Zagorski
Gloria L. Grace
Lemont E. Southworth
Russell L. Smith
Stuart E. Charlston

6. REPORT DATE June 1973	7a. TOTAL NO. OF PAGES 46	7b. NO. OF REFS 8

8a. CONTRACT OR GRANT NO. F41609-71-C-0028 b. PROJECT NO. ILIR c. ILIR-00 d. ILIR-00-31	9a. ORIGINATOR'S REPORT NUMBER(S) AFHRL-TR-73-6 9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)*

10. DISTRIBUTION STATEMENT

Approved for public release; distribution unlimited.

11. SUPPLEMENTARY NOTES None	12. SPONSORING MILITARY ACTIVITY Air Force Human Resources Laboratory Attn: AFHRL/EDA Brooks Air Force Base, Texas 78235

13. ABSTRACT

This report provides a technical analysis and review of the Computerized Academic Counseling System (CACS) designed and developed by the System Development Corporation. The system was constructed to assist counselors in guiding undergraduate college students toward the selection of optimal academic majors.

Problem review and definition, system analysis, design rationale, methodological approach, measurement specifications, data base compilation, mathematical modeling, statistical results, and validation tests are presented in various degrees of detail. Counseling application directions, capabilities, and potential are described.

Computerized academic counseling is discussed in the context of career success likelihood. Recommendations for extending the approach to include additional aspects of career guidance are made.

A concept for an Air Force career counseling system that effectively permits officers and airmen to shape their own careers is discussed. Functional components of the system include: (*a*) an Air Force personnel needs and resources forecast model, (*b*) a data base for the development and continuous support of the model, and (*c*) an Air Force mechanism which permits personnel to select careers of their choice and offers assurance that such careers will be obtained. Preliminary analyses indicate that such a system is entirely feasible and could have significant positive impact on Air Force enlistment and turnover rates. Recommendations are presented which suggest appropriate initial research and development stages.

DD FORM 1 NOV 65 1473

Unclassified

Security Classification

Unclassified
Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
counseling						
career counseling						
vocational guidance						
predicting university success						
academic grade forecasting						
university major differentiation						
computerized counseling						
automatic occupational guidance						
forecast model						
personnel turnover						
personnel recruitment						

Unclassified
Security Classification