ARTICULATION RESEARCH SYSTEM
a tool for community college curriculum evaluation
Russell Gilbert, Santa Fe Community College

ABSTRACT
Recognizing the need to evaluate the community college graduate's general education preparation for upper division coursework in the State University System of Florida, Santa Fe Community College developed the Articulation Research System. The system is currently being used by seven of the 28 public community colleges in Florida.

In determining the user requirements, it was decided that the system needed to (1) follow former community college students through their bachelor degree programs at the State Universities, (2) be able to compare student performance in community college coursework with related upper division coursework, (3) be sufficiently generalized to be able to be used by other community colleges, (4) be flexible enough to accommodate the specific student record formats at each individual community college and (5) be friendly enough to be used by curriculum evaluation administrators at the community colleges.

This paper examines how, through the use of the Statistical Analysis System, the system's requirements were met. Also discussed will be the use of a statewide course numbering system to determine related coursework, the use of a format library controlled data editing technique using a binary search and the educational impact of the Articulation Research System.

BACKGROUND
To assist in the evaluation of the effectiveness of general education curriculum revisions at Santa Fe Community College in Gainesville, Florida, several pilot projects were conducted and an early version of the Articulation Research System was developed. Funded through a Florida Department of Education computer resource sharing grant, the current version of the Articulation Research System was completed in November, 1981.

Several situations existed in the State of Florida which helped the implementation of the system requirements. First, student performance data for former community college students in the State University System (SUS) were readily available through the Division of Community Colleges. Second, the existence of a Statewide Course Numbering System enabled related coursework in the State University System and the community colleges to be identified. And last, the existence of the State University System Regional Data Centers, particularly the Northeast Regional Data Center in Gainesville, Florida, enabled the use of common software (SAS) by community colleges across the state without the need to redesign and implement the Articulation Research System for different programming languages and hardware configurations.

The Articulation Research System (ARS) was specifically developed and implemented for an IBM OS/MVS release 3.6 operating system using the JES2/MJE release 3 job entry system on an Amdahl 470 V/6-II and IBM 3033N multiprocessing computer configuration. Since the turnaround time for an information system is usually not as critical as that needed by an operational system (Martin, 1976), a batch processing system was developed to reduce processing costs. ARS was developed using the Statistical Analysis System (SAS) release 7.5 as the underlying programming language.

SYSTEM OVERVIEW
ARS Data Base
The Articulation Research System combines demographic and performance data from the State University System for former community college students each term with those students' community college history records. The combined records are stored in the ARS Data Base which has been designed in a relational data base manner (Date, 1981) in an attempt to simplify processing, to provide a simpler user view of the data and to allow modifications of the data base for individual ARS versions. The use of SAS to simulate a relational data base environment has been frequently discussed (Ingram, 1979) and debated (Bragg, 1980).

The ARS Data Base is composed of four SAS data sets:

Biographical Data Set - contains one record for each former student
Community College Course Data Set - contains one record for each course taken at the community college by each former student.

University Term Data Set - contains one record for each SUS term attended by each former student.

University Course Data Set - contains one record for each course taken at the university.

Data Independence

Since SAS separates data descriptions and decode formats from the source code programs, accommodating each college's student history record format and coding system was relatively easy to accomplish. Separate programs were developed for each college to read the community college history records into SAS data sets and edit the data elements. By adhering to certain variable and format naming and data type conventions established in the model system, the same source code programs could process data bases with different coding structures.

JCL Procedures

Nine catalogued Job Control Language procedures were developed to provide a "user friendly" environment for the ARS user. By eliminating the need to construct complex, error-prone JCL statements, the non-programming ARS user can perform all functions necessary to use and maintain the system.

As an example of the ARS JCL procedures, the following EXEC statement will create back-up copies of the ARS system libraries.

```
// EXEC ARSBACKL,
//     COLLEGE=SFCC,
//     ACCOUNT=G0001519
```

INFORMATION ACCESS

The Articulation Research System provides the user with three levels of information access:

1. Predetermined
2. Partially Predetermined
3. Not Predetermined (ad hoc).

The predetermined reports are generated by prewritten SAS programs and typically can answer only routine questions. Partially predetermined report generation is provided through the Curriculum Evaluation Product's command language which gives the ARS user the capability to ask non-routine questions within the limitations of the command language. Complete ad hoc reporting capability is provided to the ARS user by the SAS language itself. At Santa Fe Community College, this capability has been used to simulate the impact of various grading policies, to project future class scheduling needs, and to address many other "one time" questions such as the graphical comparison of student performance at the community college and the university system in figure 1 demonstrates. The user output of the Articulation Research System is discussed below in more detail.

Predetermined Output

Each time a new SUS term's data with the associated community college history data are added to the ARS Data Base, a series of user reports are generated. These reports provide the user with cross-sectional and longitudinal profile and performance descriptive statistics and student listings.

Reports which can be generated at any time by the ARS user include a combined university and community college transcript-like listing for selected former community college students, displaying biographical data for each student and each student's community college performance data, community college course records, and SUS performance data and course records for each SUS term.

The Plot/Chart Relative Term Report Series (figure 2) uses the student's relative term in the community college and university as a means of displaying longitudinal performance comparisons for many cross-sectional groups such as race, sex, program of study at the university and high school attended.

CURRICULUM EVALUATION PRODUCT

The Curriculum Evaluation Product (CEP), developed as an extension to the Articulation Research System, identifies relationships between student performance at the community college with related coursework performance in the State University System. Through a special CEP command language, the curriculum evaluator has the capability to answer an almost unlimited array of curriculum related questions.
The Curriculum Evaluation Product can identify students who have taken either a particular course or set of courses (sequence) at the community college, develop a profile of those students' performance at the State University System and compare their performance in the selected course or sequence of courses with identified related coursework performance at the university level.

To identify related coursework, the CEP user selects a set of discipline areas from the Statewide Course Numbering System taxonomy and uses their respective codes in the AREAS statement in the CEP command language (figure 3). Optionally, a magnification level can be specified which identifies the level of aggregation for the performance statistics. Magnification levels can be the selected discipline area, prefixes within the area, centuries (second digit of course number) within each prefix for the discipline area, or decades (third digit of course number) within each century for each prefix of the discipline area.

Other options available through the CEP command language include the BEGIN TERM/END TERM parameters to identify the terms in which the student must have taken the course or sequence of courses at the community college. These term parameters can be globally declared at the beginning of the set of requests or locally declared for any given course or sequence request. The CEP performance comparison statistics can optionally be aggregated by university attended and community college instructor. Figures 4 and 5 display sample CEP Profile and Performance Reports.

FIGURE 3
CURRICULUM EVALUATION COMMAND LANGUAGE
===================================
BEGIN TERM FALL 1979
END TERM WINTER 1980
COURSE BY INSTRUCTOR
MAC2311
AREAS 10(P) 68 111(C)
SEQUENCE BY UNIVERSITY
CHM2045
CHM2046
CHM2047
END
===================================

DATA EDITING TECHNIQUE

As part of the system controls, extensive data editing was employed to ensure data integrity. The State University System records contain the HEGIS discipline codes to identify university program of study, type of degree granted and the discipline of courses taken. Editing these codes presented a problem. The HEGIS codes are based on a block sequence coding structure using four digits. The first two digits represent the discipline area and the last two digits identify the specific discipline within the discipline area. Since there are approximately 500 discipline codes in the HEGIS coding system, standard range testing procedures would be of no value for validation and IF-THEN-ELSE statements were out of the question. A table look-up procedure was needed. Since the codes already existed in the ARS Format Library and had been carefully proofed and validated, a technique needed to be developed to use the Format Library in a table look-up process.

Using the FMTLIB procedure from the SAS Supplemental Library, a data set can be created with the HEGIS codes from the Format Library. Each record containing the data element to be edited can be sequentially read and, using the direct access capability of SAS, the HEGIS code data set could then be searched using a binary search algorithm to locate the matching code. Since the binary search uses an average of \(\log_2(n)-1\) comparisons (Tremblay and Sorenson, 1976) to locate an element, the average number of table accesses needed to validate a HEGIS code would be eight for the 500 values to be searched. The most comparisons needed to validate a HEGIS code would be ten (\(\log_2(n)+1\)). Figure 6 shows the source code needed to accomplish this Format Library controlled data editing technique. In actual use the algorithm has been able to validate HEGIS codes at the rate of more than 1000 records per CPU second. A real advantage of the technique besides the quick search time is that the source code is independent of the editing criteria. If the discipline codes change, only the Format Library needs to be modified and not the source code.
CONCLUSIONS

Besides providing the community college researcher with detailed graduate follow-up information, the potential to simulate, project, measure and evaluate the impact of curriculum policies and methods using the Articulation Research System creates a new dimension in curriculum planning and development.

REFERENCES


**FIGURE 4**

**ARTICULATION RESEARCH SYSTEM**  
**CURRICULUM EVALUATION PRODUCT**  
**SANTA FE COMMUNITY COLLEGE**  
**PROFILE REPORT**

* * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

* COURSE: CHM2045  
COMMUNITY COLLEGE TERMS:  
SUMMER 1978 - WINTER 1980  
* COURSE: CHM2046  
SUMMER 1978 - WINTER 1980  
* COURSE: CHM2047  
UNIVERSITY TERMS:  
FALL 1978 - SUMMER 1980  
* * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

<table>
<thead>
<tr>
<th>TOTAL POPULATION</th>
<th>71</th>
<th>2.52</th>
<th>3.09</th>
</tr>
</thead>
<tbody>
<tr>
<td>STUDENTS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AVERAGE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GPA</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**FIGURE 5**

**ARTICULATION RESEARCH SYSTEM**  
**CURRICULUM EVALUATION PRODUCT**  
**SANTA FE COMMUNITY COLLEGE**  
**PERFORMANCE REPORT**

* * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

* COURSE: MAC2311  
COMMUNITY COLLEGE TERMS:  
FALL 1977 - WINTER 1980  
* UNIVERSITY TERMS:  
FALL 1977 - SUMMER 1980  
* * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

**AREA:** 111  
**PREFIX:** QMB  
**CENTURY:** 7  
**PROGRAM:** QUANTITATIVE METHODS IN BUSINESS  
**METHODS:** IN OPERATIONS RESEARCH  
**INSTRUCTOR:** QMB-7--

<table>
<thead>
<tr>
<th># OF STUDENTS</th>
<th># OF GRADES</th>
<th>AVG SUS GRADES</th>
<th>AVG CC GRADES</th>
<th>AVG GPA</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>9</td>
<td>2.50</td>
<td>2.93</td>
<td>3.35</td>
</tr>
<tr>
<td>BRONSON J</td>
<td>1</td>
<td>3.50</td>
<td>2.96</td>
<td>3.24</td>
</tr>
<tr>
<td>COLLINS M</td>
<td>2</td>
<td>2.75</td>
<td>3.27</td>
<td>3.18</td>
</tr>
<tr>
<td>CUMMINGS H</td>
<td>1</td>
<td>3.00</td>
<td>2.94</td>
<td>3.34</td>
</tr>
<tr>
<td>DAVIDSON D</td>
<td>2</td>
<td>2.50</td>
<td>2.75</td>
<td>2.91</td>
</tr>
<tr>
<td>JONES SAM</td>
<td>1</td>
<td>0.00</td>
<td>1.86</td>
<td>3.25</td>
</tr>
<tr>
<td>ROBERTS G</td>
<td>2</td>
<td>2.75</td>
<td>3.31</td>
<td>3.90</td>
</tr>
</tbody>
</table>
**FIGURE 6**

Format Library Controlled Data Editing Technique
Using A Binary Search

```plaintext
*** CREATE EDIT TABLE ***;
PROC FMTLIB OUT=HEGIS NOPRINT;
  SELECT PGMCAT; /* GET PARTICULAR FORMAT */
PROC SORT; BY START; /* SORT BY CODE VALUE */

DATA HEGIS (KEEP=HEGIS);
SET HEGIS; /* RENAME VARIABLE */
  HEGIS=START; /* FOR READABILITY */

*** EDIT SUSCRS HEGIS CODES ***;
DATA SUSCRS (DROP=N LOWER UPPER HEGIS TRUE SEARCHIN FOUND)
  BADHEGIS (KEEP=SEN PGMCATG);

RETAIN N;
OBSNUM = 1;
SET HEGIS POINT=OBSNUM NOBS=N; /* GET TABLE SIZE */
/* SEARCH EDIT TABLE FOR EACH SUSCRS RECORD */
TRUE = 1;
DO WHILE (TRUE); /* INFINITE LOOP */
  SET IN.SUSCRS END=EOF; /* READ SUSCRS SEQUENTIALLY */
  LOWER = 1; UPPER = N; /* INITIALIZE SEARCH POINTERS */
  SEARCHIN = 1; FOUND = 0; /* INITIALIZE BOOLEAN FLAGS */
  DO WHILE ((LOWER LE UPPER) AND SEARCHIN);
    OBSNUM = FLOOR((LOWER+UPPER)/2); /* SET POINTER */
    SET HEGIS POINT=OBSNUM; /* READ EDIT TABLE */
    IF PGMCATG LT HEGIS THEN UPPER = OBSNUM-1; /* ELIMINATE HALF OF */
      ELSE IF PGMCATG GT HEGIS THEN LOWER = OBSNUM+1; /* REMAINING EDIT */
      ELSE DO; /* TABLE OR */
        FOUND = 1;
        SEARCHIN = 0;
      END;
  END;
  IF NOT FOUND THEN DO; /* PROCESS EDIT EXCEPTION */
    OUTPUT BADHEGIS;
    PGMCATG = . ;
  END;
  IF EOF THEN STOP; /* END DATA STEP WHEN */
  END; /* END OF DATA SET */
END; /* IS ENCOUNTERED */
```