Teaching the SAS® System as an Academic Subject

Andrew H. Karp
University of California, Berkeley Extension Division

Abstract
This paper discusses the author’s experiences in teaching semester-length courses in various aspects of the SAS® System at the University of California, Berkeley Extension Division. Among the topics presented are how the courses are presented, and who typically enrolls in them, key concepts and topics presented in the courses, materials developed and presented in the classes, and the anticipated future of SAS System instruction at UC Berkeley Extension.

Introduction
Many people obtain their first exposure to the SAS System in the context of an undergraduate or graduate statistical research methods course. Only when they enter the “real world” are they expected to understand and apply the SAS System’s data management and programming features to an organization’s data. Learning these skills can be a difficult task, especially if the user (or their employer) cannot afford SAS Institute courses or courses provided by “third-party” vendors. This difficulty is often more pronounced when: a) the user does not have ready access to the advice of experienced SAS System users; or b) finds their employment search frustrated by their lack of knowledge about SAS System functionalities.

This paper discusses the author’s experience in and approach taken in teaching semester-length courses in the SAS System to working adults and graduate students through the University of California, Berkeley Extension Division. As the continuing education arm of a major public research university, UCB Extension offers over 1,000 courses each academic semester at eight locations around the San Francisco Bay Area. We are one of a very few higher education institutions which provides education in the SAS System as an academic (versus corporate training) subject, and we usually exceed our maximum desired enrollment in these classes despite the presence of a SAS Institute regional training facility in San Francisco.

SAS® System Courses at UC Berkeley Extension
At present UC Berkeley Extension offers four semester-length SAS System classes:
- X401: Introduction to the SAS System
- X492: Advanced SAS System Programming
- X446: Data Analysis Using the SAS System
- X482: Advanced Data Analysis Using the SAS System.

Of these, the X401/X446 sequence is offered most semesters. The introductory class is offered at two locations in the Bay Area: San Francisco and Menlo Park. X482, Advanced Data Analysis Using the SAS System, is usually offered every third year, and the advanced SAS programming class, recently approved for inclusion in the Computers and Information Systems curriculum (see below) is tentatively planned to be offered for the first time in the fall 1996 term. Students who enroll in the San Francisco section of X401 meet twice a week for six weeks, after which Data Analysis is offered. We have offered this “two-in-one” arrangement for three years and have found that many enrollees, despite a heavy workload, find the compressed arrangement to be an effective way to learn a lot about the SAS System in a relatively short period of time.

SAS System courses each carry two semester hours of credit from the Computers and Information Systems department at UC Berkeley. They are therefore reviewed by the Academic Policy Committee of the University’s faculty Senate and are considered to be academic offerings of the University of California, Berkeley. Students may, in some instances, apply credits earned in the Extension Division SAS System classes as elective credit to graduate degree programs in other departments of the University of California. Some students have also transferred these credits to degree programs at other institutions. UC Berkeley Extension has offered courses in the SAS System for at least 10 years; the author joined the University’s adjunct faculty in 1989.

Corporate Training vs. Academic Courses
Most readers of this paper have probably attended one or more computer classes taught using “corporate training” educational techniques and are therefore able to quickly distinguish them from the academic courses they attended as undergraduate and/or graduate students. In short, the typical method by which computer training is delivered via “training mode” is different in many key respects from teaching an academic course. Among these key differences are:

- Academic Content: our courses are reviewed by the Academic Policy Committee of the UC Berkeley Faculty Senate and Instructors approved by the coordinate academic departments on the ‘main campus.’ This review and approval process means that
students enrolling in UC Berkeley Extension Division courses—regardless of academic department—are receiving an educational experience appropriate to that expected in an academic department of a prestigious research university, and that their instructor has the appropriate combination of both academic experience and (as appropriate) industry experience to teach the course material. UC Berkeley Extension Instructors are therefore usually not ‘professional trainers’ who have learned how to apply their ‘platform skills’ to a particular subject area; rather, they are subject matter experts who are able to teach the material at hand. This is, I believe, a critical distinction between ‘training’ and ‘academic courses’ which is frequently not well understood by ‘consumers’ of computer training classes, regardless of the method by which the content is delivered.

- **Duration:** Our classes meet 11 to 13 times—for two and one-half hours—per session during the semester.
- **Homework/Final Examinations:** Weekly homework assignments are a core element of how we teach our SAS System classes at UC Berkeley Extension. These homework assignments give students a truly ‘hands-on’ experience: writing a SAS System program that converts raw data in to a SAS data set and then applying BASE and/or STAT module procedures to the data sets. The in-class final examinations provide an end-of-term cumulative experience where the student demonstrates their mastery of the materials presented in the course.
- **Academic Credit:** Our enrollees may elect to take the course on one of three bases: a) audit; b) pass/no pass; or c) letter grade. Students who complete the homework examinations may take the course on a pass/no pass basis, and those who complete all of the homework assignments and the final exam take the course for a letter grade. Two semester hours of credit in Computers and Information Systems is earned by enrollees who take the class either pass/no pass or for a letter grade.
- **Computing Facilities:** although students may (and some do) use the SAS System at non-UC Berkeley Extension facilities, our students can use Version 6.04 of the SAS System at our Computer Training Facility. It is usually open six days per week, 10 hours per day. A lab tutor/teaching assistant is available one evening per week to assist students taking SAS System classes. Data sets are available on a shared drive on the CTF computing network; students using non-UCB Extension facilities are given diskettes with the data sets, and it is their responsibility to upload them to their system.

**Teaching the SAS System as an Academic Subject**

Having set out the organizational and academic contexts in which we provide SAS System instruction at UC Berkeley, the section of the paper provides more detail in how these courses are taught, including background of students typically enrolling in these courses, prerequisites, textbooks and assigned readings, and topics covered in the courses.

Our students tend to have at least a bachelor’s degree and many have one or more advanced degrees. They are therefore well acclimated to the academic environment and are highly motivated. Many are taking our classes in order to advance in their current career path or to prepare for another career direction. Some are graduate students who want to learn how to use the SAS System for an impending seminar or dissertation project. A few are retirees (either voluntary or involuntary) seeking to enhance their employment prospects.

Students who enroll in X401: Introduction to the SAS System are expected to understand basic computer concepts (such as program, file and data set), and to be comfortable using either a mainframe or personal computer. Enrolees in X446: Data Analysis Using the SAS System and X482: Advanced Data Analysis Using the SAS System must have either successfully completed X401, or have an equivalent background in using the SAS System, and must have completed at least an introductory course (for X446) in statistics and a course in multivariate analysis (for X482). Both a classroom-based and self-paced version of an introductory statistics course are offered by another department in the Extension Division.

At present I assign Frank Dilorio’s *SAS Applications Programming: A Gentle Introduction* (Duxbury, 1993) and a customized set of SAS System documentation developed via the Institute SelectText™ program for the introductory course. A SelectText™ is also used for the Data Analysis and Advanced Data Analysis courses, along with a statistics text book. The newly implemented SelectText™ service allows teachers of SAS System courses to develop, working with Institute Publications Division staff, a customized, spiral bound, text containing desired excerpts from SAS System documentation. The result is a text which is directly targeted to the specific topics that I present arranged in
the order that I present them in the class. Use of the SelectText™ program also obviates the need to make copies of SAS System documentation (with attendant concerns regarding copyright violations) or to require students to purchase several documentation manuals for the sole purpose to taking the class.

In addition to the assigned readings from the DiIorio text and the customized book, I prepare a 30 to 50 page handout for each class meeting. The handout contains a series of lecture notes, prepared using Microsoft PowerPoint, along with SAS System programs, program output, and the SASLOG. Many students like to take their notes directly on the class handouts, so I make sure they are prepared with extra-wide margins and spaces. Many students like to bind up their handouts in a loose-leaf notebook or binder, and I am gratified when students tell me that the handouts are useful to them months (and sometimes years) after they complete the course.

Instructional Method

This author utilizes a classical lecture format for presenting material in class. I try to use the same (or similar) data sets and examples throughout the same lecture period, and, if possible, across several lectures. This provides students with a chance to see how different aspects of the SAS System can be applied to the same data set, and makes it easier to show how concepts taught in one session build on those presented in previous session. Each class meeting typically starts with a review of the previous week’s homework assignment, followed by delivery of that session’s lecture. About 20 to 30 minutes at the end of the class is devoted to discussing the next homework assignment.

Homework Assignments

I believe that homework assignments are the core educational learning process for the SAS System classes I teach. The homework assignments require from three to six hours to complete, depending student ability. The assignments require students to create one or more SAS data sets from raw (usually comma-separated values) files. The assignments require that students apply programming language statements in the data step and apply one or more SAS System procedures to create printed output. Students are expected to submit both the SAS System-generated output and the SASLOG to me for review and grading. Students in the Introduction to the SAS System class are not expected to provide much in the way of data analysis since the focus of that course is on programming and data management. Homework assignments in the data analysis courses (Data Analysis Using the SAS System and Advanced Data Analysis Using the SAS System) require extensive written interpretation and analysis of the output generated by the SAS System. In those courses it is not enough to just ‘run SAS’; rather, the emphasis is on interpretation of the statistical and substantive ‘significance’ SAS System generated results.

Final Examination

Final examinations are required of students taking any of our SAS System classes for a letter grade. Common to the final examinations for SAS System classes are a series of true/false questions, short answers, code fragment correction (e.g., what’s wrong with this program?) and programming. The examinations for the data analysis courses also test the student’s knowledge of which statistical tests are supported by which SAS System procedures, appropriate characterization of the null/alternative hypotheses for these tests, and related concepts.

Students who enroll in the SAS System course sequence receive what we believe is a thorough grounding in both SAS System programming fundamental and how to implement various forms of data analyses using procedures in the BASE, STAT (and in the Advanced Data Analysis class) ETS modules. My own belief is that is much more advantageous from a pedagogical standpoint to focus in-depth on a smaller number of concepts that to spend relatively shorter periods of time on a larger number of topics.

The sheer magnitude of SAS System capabilities makes presenting classes about it an interesting challenge: there is so much one could teach, and so little time to teach it. Also, the varied academic and occupational backgrounds of students enrolling in UC Berkeley Extension SAS System classes (and all of our Computers and Information Systems classes), coupled with high motivation levels, often means that I am challenged by my students to address topics not contemplated by the course syllabus. The result is lively and active class sessions punctuated by questions and ideas from students.

Course Objectives

One common method by which to organize and plan a course is to develop a set of objectives. These objectives can be stated in the context of a certain body of knowledge that the student will acquire or a set of tasks/skills they will be able to perform at the end of the term. I’ve found this to be an effective way to set up my SAS System classes at UC Berkeley Extension.

This has particularly been the case for the Introduction to the SAS System class. For that class I have established the following objectives for student learning:

- Convert raw data to a SAS data set using the INPUT statement.
• Combine SAS data sets using, as appropriate, merging, concatenation, interleaveing and subsetting
• Perform conditional, arithmetic and statistical operations on the values of variables in a SAS data set, including use of SAS functions
• Summarize and report the values in a SAS data set using BASE SAS procedures, including PRINT, FREQ, TABULATE and SUMMARY
• Use the data visualization procedures in the BASE module, such as CHART and PLOT
• Obtain the descriptor portion of a SAS data set using PROC CONTENTS
• Modify the descriptor portion of a SAS data set using PROC DATASETS, as well as use the procedure to perform ‘housekeeping’ operations on a SAS data library
• Re-arrange the contents of a SAS data set using PROC SORT
• Modify the external representation of the values of variables in a SAS data set using both internal SAS formats in the BASE module as well as those the user creates using PROC FORMAT
• Understand how the statements in a SAS data step are executed in the Program Data Vector (PDV), and from this understanding, obtain insights into SAS System programming efficiency techniques which reduce storage requirements and program execution time.

The course objectives for the data analysis courses are much more straightforward and tied to the specific subject areas covered, such as correlation, analysis of variance, regression, principal components analysis and the like. For the data analysis courses I motivate the concepts presented in the classes using the well-known ‘Steven’s Taxonomy,’ which classifies variables as being nominal, ordinal, interval or ratio. The classification of variables in this fashion permits rapid identification of statistical tests appropriate to the analytic situation at hand.

Each class session, homework assignment and the final exam are all tied to one or more of these objectives. Announcing these objectives during the first class session helps students understand the anticipated course content and helps them decide whether the course will meet their needs. This is particularly important in learning environments like UC Berkeley Extension, where no formal admissions standards are imposed; rather, students ‘self-select’ themselves for enrollment in our courses.

A Unifying Theme: The ‘Prime Directives’ of the SAS System
The scope and breadth of SAS System functionalities makes teaching a course like Introduction to the SAS System a challenge. What to present? What to leave out? How can the core concepts underlying use of the SAS System be taught? One answer to the last question was to develop a set or core, or condensing, statements which I believe make it easy for users to grasp the fundamental concepts of using the SAS System. For woe of a better term I have come to call them the ‘Prime Directives’ of the SAS System. My experience is that if a student in my Introduction to the SAS System class can remember the following ten things, he or she can learn to use the SAS System:

1. all SAS System programming statements end with a semicolon
2. the maximum length of a SAS data set or variable name is eight characters, and it must start with a letter or underscore
3. a SAS System Unit of Work is terminated by either the SAS programming statement RUN; or by another Unit of Work
4. Units of Work are either Data Steps or Procedure Steps. and a single SAS program usually contains more than one Unit of Work
5. the SAS System operates on Units of Work in the exact order that they appear in the program
6. the Data Step operates on one observation at a time
7. a Procedure operates on all observations in a Data Set
8. avoid using Data Step statements in a Procedure Step, and vice versa
9. there are at least three ways to do any one thing in the SAS System: one Wrong Way and at least two Right Ways. One Right Way is always more efficient that other Right Ways
10. for every prime directive there exists at least one exception, knowledge of which is not required to learn the fundamentals of the SAS System.

These ‘prime directives’ were the result of several informal conversations I had with friends in the Bay Area (California) SAS User’s Group. I had initially anticipated that this list would serve as something of a humorous introduction to the course. As it turned out, they were an unusually effective way to condense key

1674
concepts in SAS System usage and explain them to new users. One method by which I apply the prime directives is to constantly refer to them in my lectures. I might, for example, answer a student's question by starting with: "well, as prime directive #5 states, the SAS System operates on Units of Work in the exact order they appear in the program, so we would need to first do a data step to create the data set we need before using PROC FREQ to obtain a frequency table..." In fact, I have even graded homework assignments by reference to the 'prime directives.'

Future Directions for SAS System Instruction at UC Berkeley Extension
As noted at the outset of this paper, UC Berkeley Extension has offered courses in the SAS System for about a decade. The Data Analysis/Advanced Data Analysis courses were added in 1989 and the Advanced SAS System Programming course was approved in 1995.

Class enrollment usually exceeds our maximum desired enrollment level of 33 attendees per Computers and Information Systems semester-length course. We maintain these enrollment levels even though SAS Institute has begun to offer an array of two- and three-day courses at their newly opened SAS Institute regional office. Some people may apparently prefer, and have the financial resources, to attend a corporate training session while others find the academic class format more suitable to their needs. That both are able to co-exist in the same geographic market area suggests that there is an ongoing need for both forms of instruction in the SAS System.

Many of our students in SAS System classes have asked us to provide additional, more advanced courses. In addition, we perceive the need for an organized, formal course of study in the SAS System, upon completion of which a student would receive some form of academic recognition from the University of California.

The Proposed Certificate Program in SAS System Studies at UC Berkeley Extension
Addition of the Advanced SAS Programming class is just the first step towards meeting those needs. Earlier this year UC Berkeley Extension initiated the processes to start a Certificate Program in SAS System Studies. We believe that our proposed SAS System certificate program will be the first such professional certificate program in the SAS System to be offered by a college or university in North America.

While not a 'degree,' we hope that our anticipated SAS System Studies certificate will provide enrollees with a rigorous set of courses in the SAS System and related topics, completion of which will have some level of academic and/or professional recognition. At present, the proposed program will be comprised of a total of eight, two-semester-unit, classes, of which six will be in SAS System-related classes.

As presently proposed, the Certificate in SAS System Studies will include six courses from the following:
- Introduction to the SAS System
- Advanced SAS System Programming
- Data Analysis Using the SAS System
- Advanced Data Analysis Using the SAS System
- Data Visualization Using the SAS System (new)
- Applications Development Using the SAS System (new)
- Report Writing Using the SAS System (new)
- SAS System Macro Language (new)
- SAS for Forecasting and Time Series Analysis (new)

Of these, Introduction to the SAS System would be required or could be waived via challenge examination.

One of the remaining two classes will be a course in computer systems consulting and the other will be in a programming or data base language, to be chosen from our other Computers and Information Systems classes.

We anticipate a decision about our proposed SAS System Studies certificate about the time SUGI 21 is held in March, 1996.

Conclusion
Teaching the SAS System as an academic subject is appropriate for adult students and others who wish to learn about SAS System functionalities in a traditional academic format of lectures, homework assignments and examinations. Development of a set of objectives for each course, as well as course materials (both by the instructor and using SAS Institute’s new Selcetext™ program) for student use, is essential for presenting the material in a logical and effective manner. Further, use of the 'prime directives' helps as a tool to unify the principles taught in the class. The proposed Certificate Program in SAS System Studies was developed to meet students’ needs for a formal course of studies in the subject matter area.

The author can be contacted at:
1489 Webster Street #1308
San Francisco, California 94115
415/441-0702
SFBAY0001 @ AOL.COM

SAS is a registered trademark of SAS Institute Inc. in the United States and in other countries. ® indicates U.S. registration.