

# THE OBSOLESCENCE OF THE SAS SYSTEM

GRANT BLANK, Independent Consultant

## ABSTRACT

The SAS® System has established itself as one of the standards against which quality software is measured. It is so good that, in the area where it began, statistical analysis, it has virtually no direct competition. Yet, the environment for the software that SAS Institute provides is changing. Users are becoming more knowledgeable and are demanding more sophisticated but easier-to-use information processing and data analysis capabilities. Furthermore, microcomputers provide a much more user-friendly environment than mainframes and their speed and memory capacity are increasing rapidly while their costs are dropping. The software appropriate for this environment is different than that designed for other environments. Its hallmark is a user interface that is highly interactive, easy-to-learn and easy-to-use. The SAS System is not well-adapted to this environment. If the SAS System is to retain its current pre-eminence SAS Institute will need to make major changes in its syntax, in the ease with which data can be manipulated, retrieved and stored, in its documentation, and in the ways in which results from statistical processing can be used.

## INTRODUCTION

When people or organizations pass certain ages it is customary to evaluate their life; where they are, whence they have come and where they are going. As this SAS User's Group begins its second decade, it is appropriate for us as users to examine the SAS System. We need to ask: How has it served our needs? How can it continue to serve them in the future?

My approach to these questions has five parts. The paper begins by discussing the factors that led to the success of the Statistical Analysis System during the 1970's. The second part of the paper surveys the environment in which software like the SAS System must compete today, in the mid-1980's. Part three suggests that the SAS System is not as good a product for the 1980's as it was for the environment of the 1970's; that the SAS System, in its current form, may be obsolescent. This leads to suggestions for changes needed to make the SAS System the superior product that it was 10 years ago. The paper

concludes by identifying the fundamental characteristic of superior software and suggests incorporating this characteristic throughout the SAS System should be the primary goal of SAS Institute in the coming years.

The conclusions that I reach are personal views. But I have a great deal of experience with the SAS System. I have been a SAS user since 1977 and I continue to use the SAS System, practically on a daily basis. Furthermore, as an Independent Consultant, my experiences have been both broad and deep. I have contact with, and can speak to, the interests and needs of a wide variety of users.

## WHY WAS THE STATISTICAL ANALYSIS SYSTEM SUCCESSFUL?

To understand why the SAS System has been so successful, we need to begin in the mid-1970's when the Statistical Analysis System first became a commercial product. When I encountered SAS in 1977 I, like many other people, quickly became an avid SAS user; some people said I was a SAS fanatic. What was it about SAS that made people like me think it was so wonderful?

In the middle 1970's if you wanted statistics you had three major choices: SAS, SPSS® and BMDP®. SPSS (Release 7) dominated the market. It was a simple little package that couldn't do very much and many of the statistical things that it could do, were not done well. Its strongest point was its simplicity--anyone could learn to use it quickly and easily. This strength was enhanced by an excellent tutorial user manual.

The problem with SPSS was that it couldn't handle anything that wasn't simple. Among its many problems the most serious are: severe, and poorly explained, restrictions on the ordering of statements; arbitrary restrictions on the capability of most statements (e.g. only a certain number of variables could be labeled on a single VAR LABELS card); incomprehensible error messages when you exceeded these and other limits; non-numeric variables restricted only to FORTRAN A4 format; and severely limited file handling capabilities (e.g. limited ability to merge files and no ability to concatenate files). To overcome these limitations, those of us who used SPSS spent much of our time in other languages, manipulating the

data to get it into a form that SPSS could handle. As a consultant, I made a lot of money this way but it was slow, tedious and clumsy.

BMDP, on the other hand, had excellent statistical capabilities. Its statistics were not only better than SPSS's, they were better than SAS's. It had a simple, free-format command language, with few of SPSS's restrictions. In other areas, BMDP shared SPSS's weaknesses. Its ability to handle character variables was as inadequate as SPSS's. Its error messages were, if anything, even more cryptic. But BMDP's greatest weakness was that it made no pretense of having any data transformation or file handling capability. Even the simplest transformation required using FORTRAN. Like SPSS, using BMDP required spending a lot of time in other languages, tediously manipulating data.

Why the SAS System replaced SPSS as the dominant statistical package is easy to see: SAS could do almost everything better. But BMDP is different. There is a paradox here. Why did a package calling itself the Statistical Analysis System come to dominate the market when a competing package produced better statistics?

The SAS System distinguished itself from its competition in the middle of the 1970's for five reasons. These are really the reasons why I became a SAS fanatic.

1. SAS's powerful file handling capabilities were available nowhere else. With MERGE, SET and UPDATE, I could complete, in minutes, file manipulation problems that would have stretched on for hours or even days in any other system.
2. The SAS System had an extremely flexible approach to reading raw data. I could read almost any file format and almost any kind of data.
3. The programming capabilities of the Data step were strong enough that practically anything I wanted to do in PL/I, I could do more easily and faster in SAS.
4. SAS's statistical procedures were usually adequate. When they weren't, PROC BMDP provided an easy interface to BMDP's excellent statistical capabilities. I could use the SAS System to manipulate and transform my data and then, when I needed complex, multivariate statistical procedures, run them in BMDP through SAS.
5. To be fair to SAS's statistical capabilities, there was one exception to the rule of

mediocrity. While other SAS statistical procedures were undistinguished, PROC GLM was different. It was a creative and innovative product.

Combining the ability to read a wide variety of raw data formats with extensive file handling, adequate statistics, an interface to excellent statistics and reasonable programming capabilities gave me the option of staying inside the SAS System to do almost anything. And I had access to all this through a single integrated system with a single, consistent syntax, a single file type and only two data types. The SAS System was enormously **flexible**. I needed better statistics than SPSS, which were widely known to be inaccurate.[1] But I didn't need outstanding statistics nearly as much as I needed to get away from using FORTRAN and PL/I for file manipulation and transformation. Using the SAS System **simplified** my life.[2]

That is what made me a SAS fanatic: The combination of **simplicity** and **flexibility**. The case of BMDP makes clear that only being statistically superior was not enough. Both BMDP and SPSS were simple, but they weren't flexible. This was the failure that led to their decline. Simplicity and flexibility are not easy to combine in the same software. They exist in tension with each other because they impose partially contradictory requirements. SAS Institute's success can be directly attributed to its ability to resolve that contradiction into a coherent, even elegant system in the context of the batch-processing environment of the 1970's.

Since the 1970's the SAS System has become the standard for statistics and graphics on mainframes. Everyone compares their software to SAS products. On mainframes today, the SAS System is so dominant that it has only minor competition. Its competition is mostly in specialized areas of statistics or graphics; or is restricted to non-IBM hardware. SAS Institute has been so successful that in the last several years it has begun to expand into non-statistical software. Symbolically, it has changed its name from the Statistical Analysis System to the SAS System.

## TRENDS IN HARDWARE AND SOFTWARE

Now we're in the middle of the 1980's. In the world of computers, that's a long way from the mid-70's. The environment in which we use SAS products has changed in many ways. Before I examine what has happened to the SAS System, let me review how the environment is changing.

**Hardware Developments.** The direction of hardware development is clear: very rapid increases in speed, memory and storage at lower

cost. Further developments include multiple windows on much larger screens. The net effect will be to make much more computing accessible at lower cost. One important implication is that the market for computing services will continue to widen. These changes are the foundation of most other developments

**New Applications.** As costs fall it has become cost-effective to use computers to solve many additional problems that are unique to a single organization: for special report formats, unusual datasets or analyses, or other special purposes. Meeting these needs requires the ability to develop applications not included in off-the-shelf software. The SAS System can be used to develop new applications, but it is not well adapted.

**Data Base Management Systems.** There are several very good DBMS's being marketed today. The SAS System is not a DBMS, it is a package designed for statistics, graphics and report writing with file handling capabilities. You may think that the two wouldn't compete with each other. But vendors of DBMS's are discovering that their customers not only need to print reports, but they want to do statistical or graphical processing on their data before it is put into a report. So they're adding statistical and graphical capabilities and we're beginning to see more sophisticated statistical capabilities, much like those in the SAS System. Furthermore, for many purposes, a DBMS is easier to use than a file handler.

**Microcomputers.** Today virtually all new users receive their introduction to computing on microcomputers. Micros can't compete with mainframes or minicomputers for most production computing. They don't have the speed or capacity. But microcomputer capabilities are increasing quickly and most new computer applications don't require the speed or storage capacity of a mainframe. New users are being introduced to computing on spreadsheets and word processors. These are fundamentally different than the sort of programs that were available only a few years ago.

A key characteristic of micros is their highly interactive, user-friendly terminal interface. User-friendly systems are easy to learn and easy to work with, so people don't need much training to become productive. The best of the current set of microcomputer software, for example Lotus 1-2-3, have such a clean design that many new users can do sophisticated substantive work within a couple of hours of being introduced to the system. Nowhere in the SAS System is there anything that can match this kind of friendly, capable environment. The easy-to-use, interactive user interface of a microcomputer is a stark contrast to

the user hostility of many mainframe products like JCL or IMS-DL/I.

**User Expectations.** Users have become much more sophisticated about what they can expect from software. I find, increasingly, that clients who have experienced the benefits of well-designed microcomputer software are very reluctant to accept software which doesn't provide these user-friendly features, regardless of its other virtues. People like this won't even look at a product like the SAS System.

**User Needs.** As the market for computer-processed information has widened, the base of experienced users has also broadened. With cost reductions and increasing speed, more and more people have cost-effective access to statistics. Users of low-cost computing power will have different needs than the people for whom it was cost-effective to use computers just a few years ago. This only reinforces the trends discussed above. Since they didn't pay very much for their computer, they don't have any trained computer people on their staff. They don't want to hire computer professionals, they want to do it themselves. Or give it to their secretary or assistant to do. These trends only add to the pressure that makes **ease-of-use** the most important single characteristic of software.

## IMPLICATIONS FOR THE SAS SYSTEM

The environment in which the SAS System was developed is very different than the environment I have just described. Today people compare the SAS System, not with other statistical packages, but with DBMS's, spreadsheets and word processors. SAS Institute products don't compare very well. They are not easy to use compared to spreadsheets and they're not sophisticated in the ways that a DBMS is. The point is: As the market has expanded to new groups and new competitors, the SAS System has not kept pace. In very fundamental terms, the old synthesis of simplicity and flexibility that was the foundation of SAS's pre-eminence as a statistical package in the 1970's, has broken down in the 1980's and it has not been replaced.

The SAS System has become much more complex than it was. Features have been piled on top of features. Instead of the clean, simple package it was in the 1970's, SAS has become bulky and complicated. In a curious sense, the SAS System has also become less flexible. Users in the 1970's would remark, happily, how easy their project could be done in SAS compared to SPSS. Today they are more apt to wish that Lotus Development Corp. (the developers of Lotus 1-2-3) had written SAS

because they surely would have written more flexible software.

The issue is not, of course, that the SAS System has become less flexible, rather new users, new hardware and new software have **fundamentally changed** the market in which SAS competes. The standards by which the SAS System is judged have altered. SAS has also changed, but the changes have been **incremental** rather than **fundamental**. These changes will not be sufficient to keep the SAS System competitive. The next release of the SAS System (Release 6) must be much more than another incremental change. SAS must become a very different system.

## WHAT CAN BE DONE?

Here are some of the important changes that I look for in SAS Release 6:

1. SAS doesn't have any decent interactive system. The existing interactive SAS is basically batch-SAS, thinly reworked so that a user can type commands at a terminal. The Display Manager is a useful advance, and if it had been available in SAS82, it would have been a great product. In the intervening years other products have become available and the Display Manager doesn't compare well. It's not as friendly, its error-correction capabilities are not as good, it requires substantial computer resources and it forces users to learn yet another editor. The SAS System is not competitive with a truly interactive system.
  - The syntax of the commands is wordy. It was designed for batch use, not as an interactive system.
  - As the SAS System has grown, inconsistencies have crept in. Compare, for example, the output options in FREQ and SUMMARY. Some procedures will not accept statements in any order (e.g. SUMMARY and GMAP). Such inconsistencies force the user to learn arbitrary, special rules, seriously compromising the simplicity that was once one of the SAS System's greatest strengths. I expect that they also indicate management and control problems inside of SAS Institute.
  - Help files are small and limited. They need to be expanded and made available everywhere for every statement and option in SAS. Currently, the help files are not context-sensitive. This makes them harder to use.

- Error recovery is very weak. Errors cannot be fixed interactively, except in very clumsy ways. The Display Manager simplifies error correction but there is still no way to diagnose and correct errors line-by-line as they are entered. Entire Procedure and Data steps have to be rerun.
- Error checking during data input is difficult. FSEDIT is particularly user-hostile in this respect. Since it updates in place, an error on data entry can actually damage a file. The new SAS/FSP Procedure, BUILD, will have much better error checking. This needs to be extended across the entire SAS product line.

2. Documentation has always been the weakest part of the SAS System. All users depend heavily on written documentation and the documentation has never been as good as the software. Here are some of the worst problems that need correction in Release 6.

- Documentation is almost solely restricted to reference manuals. Except for a tiny beginners guide, there is no easy way for a new user to begin to learn the system. The classes that SAS Institute and others give are not effective substitutes for a good set of tutorial manuals. Video training is another excellent idea but it is still not a substitute for a good set of tutorial manuals. The point here is that no matter how a training program is organized, there is still a very long gap between the end of the training and reasonable acquaintance with the 2000-odd pages in the Basics and Statistics User's Guides. Tutorial manuals are badly needed.
- The state of the documentation is an excellent example of how the underlying coherence of the SAS System has broken down. Looking at the 2000 pages the user sees only complexity, the underlying simplicity of the SAS System is completely lost in the morass of details. The User's Guides offer little guidance as to what is fundamental or general and what is of esoteric interest.
- One result of the poor documentation is that users ask for much more technical support from SAS Institute. This is a drain on the resources of the Institute, and probably a management headache as well.
- Some sort of interactive, computer-aided instruction available within the SAS System would be extremely valuable. In its most useful form, this would be a tutorial facility

available within the Data step and all procedures.

3. The SAS System needs to be much more easily extensible. Look at modern languages like FORTH, Modula 2, or C. They are marked by an easy ability to insert new capabilities into the language by using the language itself to write a new module.
  - The Macro language is a good idea with a fundamental flaw: It is similar to basic SAS but subtly different in many ways. This means that use of the macro language requires learning a new language. The learning process is complicated by the fact that the language is close enough to SAS that users find it difficult to keep track of important distinctions. The language is different from SAS; it is poorly documented; and not very readable. As a result, it is inaccessible to most users. In my opinion, the macro language needs to be simplified and made more accessible. Someone needs to do for the macro language what SAS did for PL/I.
  - In general, the interface to user-written procedures is not suitable because it requires much special knowledge about another language (e.g. PL/I) and various specialized software (e.g. linkers). This means that just to get an application up and running requires fairly heavy front end costs. Most SAS users do not have the ability to use this facility, even if they wanted to. Furthermore, being forced into another language again undermines the simplicity of the SAS System.
4. It is very difficult to get information that is printed on SAS output pages into a SAS dataset. Why can't results from statistical calculations be much more readily available? The SAS System would benefit from much more memory-to-memory data transfer.
5. File handling needs to be much more flexible and easy. Particularly, users should not have to learn relational database management theory in order to use the SAS System. But now they have to.
6. The SAS Data step was a fine file handler for the 1970's. It is adequate for simple data retrieval and reporting. Today, users need much more complex database management and the SAS System is inadequate. It badly needs an interactive query language. This means adding capabilities to handle multiple write access and inverted lists to SAS datasets. These are not

trivial tasks, but they are extremely important and they are well within the capability of SAS Institute. The acquisition, last December, of the System 2000 data base management system is a very hopeful sign that SAS Institute understands this problem. I look forward to being able to use it on SAS databases using SAS-style syntax.

These weaknesses are interrelated but can be summarized. Specifically, there are four major problems:

- I. Limited interactive capabilities;
- II. Serious flaws in documentation;
- III. Limited ability to extend SAS's capabilities; and
- IV. Lack of DBMS capabilities.

These problems were not so serious when the only alternatives to the SAS System were little SPSS and BMDP. They are much more important today.

## CONCLUSION

AT&T is marketing a statistical command language called 'S'.<sup>[3]</sup> It will allow much more powerful data manipulations than any other statistical language. AT&T and others are working on an expert system which will make use of that language for statistics and graphics.<sup>[4]</sup> It will incorporate capabilities that will allow it to examine the data and suggest analyses. For example, if you were to input a regression model it may respond, "The residuals are curved, may I suggest a log or a square root transformation?" If you were looking at a scatterplot of, say, trends in sales growth, this system might say, "Trends are often easiest to see by looking at period-to-period differences, may I suggest a plot of first differences?" This software is currently being developed for research purposes. But, in an environment where there are lots of small computers and many people who need statistics but don't have very much statistical training, the commercial potential of a statistical expert system is immense.

Expert systems are a promising new direction. Imagine how much simpler information processing would be if the software itself suggested fruitful ways in which the data could be analyzed or graphed. Within five years, I expect commercial products will be marketed that will enable you and I to produce better statistical and graphical analyses than most consulting statisticians can do today. Is SAS Institute conducting research on expert systems? I don't know the answer to that question, but I don't see any evidence of it here at the SUGI conference.

When SAS was introduced as a commercial product in 1976, the dominant statistical package of the time, SPSS, looked at capabilities such as MERGE, the ability to sort a dataset freely, at SAS's extremely flexible data input capabilities. When I talked with the SPSS staffers they said things like, "Nobody needs these things. They don't even have a RECODE statement. Why should we worry?" As a result of such complacency, among other weaknesses, SPSS is no longer a serious competitor and it continues to lose market share year by year.

The computer software business is a competitive, rapidly-changing environment. The company that thrives in that environment will be characterized by an openness to the demands of the environment. It must be flexible, ready to change its products and itself as the environment requires. Finally, it must be aggressively managed. To me, this means it must identify and design the sort of products that many of its users may not yet realize that they need to do their tasks productively and efficiently.

That, ultimately, must be SAS's goal. It may sound surprising, but I do not think that SAS Institute is in the computer software business. Its business is making people productive and efficient. To remain competitive it must once again recapture what it once did well. It must design simple, flexible software. This is characteristic of software that simplifies people lives and makes their problems easier to solve.

The SAS System is already substantially less than it could be. It has lost the simplicity and flexibility that were the foundation of its initial success. I would like for SAS Institute to give me innovative software so that I could be as much of a SAS partisan today as I was in the late 1970's. But the recent additions to its product line such as those in SAS/FSP, the Interactive Matrix Language, the Display Manager or SAS's spreadsheet, PROC FSCALC, aren't very exciting, and they clutter rather than simplify. They are not the sort of innovative, simplifying software that SAS Institute provided with GLM or the ability to do relational database management with MERGE and SET. They are not products that lead the way into new directions.

SAS Institute is burdened with an aging product line. It has not kept pace with current developments in software design. If it doesn't respond to the needs of its users and to technological reality, it will become the has-been company that some of its competitors are now.

#### FOOTNOTES

[1] There are many anecdotes about this. In my case, an SPSS (Release 7) factor analysis once informed me that it explained over 300% of the variance. When I saw the output, I laughed, showed it around the office and threw it away.

Now I wish I had kept it. For a more scientific approach, see Wilkinson and Dallal, 1977, and Velleman and Welsch, 1981.

[2] Two examples can help clarify what I mean by simplification. The first example is the capabilities of the MERGE and SET statements and the executable FILE statement. Anyone who has tried to work with datasets without using relational database operators such as these, knows that it is a slow and difficult process. Using the SAS System, MERGEs and other relational operations become simple, straightforward operations that even novices can use successfully. Example two is PROC GLM. GLM simplifies statistical work by giving user's the ability to approach a group of related statistical topics using a single, consistent approach and syntax. Of course, GLM won't simplify the statistics but it does simplify the process of instructing the computer. In both examples, the SAS System provides tools that dramatically simplify complex tasks.

[3] See Becker and Chambers 1983 for details on S.

[4] For a general discussion of statistical expert systems and a description of the AT&T project, see Hahn, 1985.

#### ACKNOWLEDGMENTS

I would like to thank the following people for their comments on earlier drafts of this paper: Rebecca Blank, Ed Brent, James Campbell, J. Michael Camp, Robert Cross, Joanne DiPietro, Christian P. Gruber, Robert Lerner, Linda Muthen, Andrew A. Norton, J. Robert Passmore, Nora Schaeffer, Andrew Shapiro and Leland Wilkinson. Inasmuch as I did not always accept their advice, I am responsible for any remaining errors or omissions.

#### REFERENCES

- Becker, R. A. and J. M. Chambers,  
1983 S, An Interactive Environment for Statistics and Graphics, New York: Wadsworth.
- Hahn, Gerald J.,  
1985 "More Intelligent Statistical Software and Statistical Expert Systems: Future Directions," American Statistician, 39: 1-6.
- Velleman, Paul F. and Roy E. Welsch,  
1981 "Efficient Computing of Regression Diagnostics," American Statistician, 35: 234-242.
- Wilkinson, Leland and Gerald E. Dallal,  
1977 "Accuracy of Sample Moment Calculations Among Widely Used Statistical Programs," American Statistician, 31: 128-131.

SAS and SAS/FSP are registered trademarks of SAS Institute Inc., Cary, NC, USA.

SPSS is a registered trademark of SPSS, Inc. Please note that all references to SPSS in this paper refer to the SPSS batch processing system (Release 7) in use in the mid-1970's. At no point do I refer to more recent products of SPSS, Inc.

BMDP is a registered trademark of BMDP Statistical Software. All references to BMDP refer to the software available in the mid-1970's, not to current or planned BMDP products.

Comments or criticisms are welcome and solicited.

Grant Blank  
Independent Consultant  
1450 East 55th Place, Apt. 922  
Chicago, IL 60637