Novel Approaches to Learning and Teaching SAS-based Analytics

Richard J Self BA(Cantab) LLM MBCS and D Voorhis BSc MBCS
School of Computing and Mathematics
University of Derby
r.j.self@derby.ac.uk and d.voorhis@derby.ac.uk

Abstract

Traditional approaches to introducing students to computer programming typically involve carefully constructed teaching lectures and workshop exercises using programming languages like Java, C# or Python. As a general rule, this works for some groups of students. However, many less technical – but no less intellectually able – students find the approach intimidating and fail to develop any productive skill in the taught language.

The approach presented in this paper is fundamentally different and was used to enthuse a group of Information Technology students at the University of Derby (UK) to learn Base SAS. The majority of the students had failed to successfully engage in technical programming in their previous semesters and many provided strong indications that they did not want to be involved in any form of programming.

The approach described in this paper was based around the tried and tested pedagogic approach of “learning by doing”, but modified to become “Learning by Exploring”. The results of this approach were highly positive, resulting in a 98% pass rate, with 77% of the students gaining an overall mark of 60% or higher for the programming and analytics part of the assessment. In addition, six of the forty three students researched, found, learned and used constructs of the provided SAS environment that were unexpected and would have never been taught or learned in a traditional approach but were highly innovative and appropriate solutions to the defined task.

Introduction

The School of Computing and Mathematics at the University of Derby has recently become a member of the SAS Student Academy network in the UK and is starting the process of embedding SAS in the curriculum for the BSc (Hons) Information Technology programme. A key driver of this initiative is the growing significance and importance of Big Data and business analytics / business intelligence in the business environment where our graduates will become employed. It will also provide a clear and distinctive employability boost for our graduates as we deliver the opportunities for them to gain SAS Certification.

The majority of the students on the BSc (Hons) Information Technology programme choose the programme because of the emphasis on the use of Information and Communication Technology (ICT) rather than on the technical development and application of technology in programmes that cover computer science, computer networks and security, or computer forensics. Information Technology students are more interested in the Architect / Designer levels of the Zachman Framework for Enterprise Architecture (Zachman 2013) compared to computer science students who are generally much more interested in the lower, technical
levels of the Zachman Framework. As a result, the majority of the students have little interest in learning any programming language; many of them had minimal ability even with Microsoft Excel. This clearly posed a significant challenge to the School of Computing and Mathematics in terms of overcoming this attitude.

It was clear that the traditional, highly-structured approach to teaching programming languages would not be appropriate for this cohort of students. As a result, a novel approach was developed that enticed them into using the environment as an easy way to gain answers from sets of data (small rather than Big) in a second semester, second year module called *Data Management and Business Intelligence*.

**Teaching Context**

The students had previously (autumn semester of 2012) taken three modules which could be considered as foundations for this module: *IT Services Management, IT Product Design* and *Databases*.

The module was specified to use SAS as the vehicle to explore the academic and technical aspects of the topic, due to the experience of two faculty members (the authors) during the 1980s and 1990s with SAS in the business world as an environment for analytics and a tool for deploying end-user computing.

The module was taught through seminars and workshops with four hours of contact time per week in groups of 20 students in dedicated computer labs. The seminars were designed to lead the students to research – and learn – the theoretical underpinnings of the subject. The workshops were designed to allow the students to explore and apply the concepts using SAS installed on the lab PCs, and on the students’ laptops under the academic licence.

The assessment was to develop a portfolio comprising two parts: a short requirements’ specification followed by the creation of a small Management Information System (MIS), implemented in SAS, based on a defined set of public transparency data from the UK National Health Service (NHS 2013).

The requirements’ specification was required to define and justify five aspects of the required MIS. It was limited to between 1000 and 1500 words to represent a typical user specification for a small application in a business context.

The MIS was required to host the data from a number of the monthly files that originated from the data source, and then provide three simple reports (using graphs, charts, tables, etc.) that would answer questions relevant to one of the data stakeholders – which could range from hospital patients up through the NHS hierarchy to government ministers. A rubric was provided to guide the planning and evaluation of the artefact that would be demonstrated for the assessment of the MIS. The rubric identified the different mark levels to be awarded to different levels of complexity and technical prowess. Two factors were assessed (Coding and use of SAS, and Data Structures). A key feature of the first factor was that the highest levels of marks would be reserved for a fully menu-driven system.

**Approach**

Based on the need for a different approach to gaining the students’ interest in developing code for the Base SAS product, it was decided to provide some simple tutorial materials on
how SAS worked. Research suggested that the materials developed by the Institute for Digital Research and Education at UCLA (UCLA 2013) could provide an interesting set of materials for self-study and exploration of the ease of use and simplicity of coding and production of useful reporting in tabular and graphical terms. This material was used during the workshops and in their private time. In addition, the students where shown where a range of publically available SAS manuals could be found covering, a range of the Base SAS product set, with the suggestion that these manuals could be useful.

The workshops were organised to provide opportunities for the students to work through the UCLA material and provided demonstrations of the need for, and examples of, the exploration of data that is required before developing any analytics activity.

The demonstrations of the approach to exploring the data were carried out in Microsoft Excel because the source data was provided in that format and it demonstrated the necessary pragmatic approach, and because the students were generally familiar with Excel – albeit with widely varying skill levels. It also turned out that the actual Excel spreadsheets contained various data processing challenges that needed to be understood before importing the data into SAS.

The workshops also included refresher material covering normalisation of data and illustrations of the types of 3rd normal and 6th normal forms that they might consider using in their small MIS products (also associated with the marking rubric requirements).

The specification of the portfolio primarily concentrated on the requirements’ capture process as a detective experience, based on the experience of the authors in their pre-academic career. The specification of the SAS artefact was provide solely by the assessment rubric, see below for an example of the top level definition.

<table>
<thead>
<tr>
<th>Percentage Band</th>
<th>Coding and use of SAS</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>90%-100%</td>
<td>Relevant SAS Source code pre-written and stored in SAS system. Code modules chosen and run from a menu system. Data selection menu driven with complex options. Graphic data presentation, good choice of output styles which are menu driven.</td>
<td>Uses pre-imported data in a saved master file using the relevant data for 2011 and 2012. Optimises SAS storage structures to the types of questions provided. Optimises storage space by use of surrogate keys to natural key tables. Optimises use of permanent and work file structures.</td>
</tr>
</tbody>
</table>

The students were offered the opportunity, for those gaining 70% or more, to undertake a boot camp leading to the Base Programming Certification test, to be paid for by the School.

An additional incentive was provided to the students with the provision of a link to a UK website which identifies salary levels for a range of jobs and required skill sets (ITJobs Watch 2013) in order to identify the relevance SAS and thereby generate enthusiasm.

Particularly notable is that students were not given prescribed steps to completing the assessment, but were explicitly encouraged to explore the SAS tools on their own. Thus, this “learn by exploring” experiential learning (Kolb 1984) variant of “learn by doing” explicitly employs elements of problem-based learning (Hmelo-Silver 2004) and enquiry-based learning (Edelson et al. 1999).
Results

Provision of the ITJobs Watch link – which indicated at least £10K salary increment compared to normal student salary levels for our graduates – generated significant enthusiasm of many students, as shown by the level of activity both in the workshops and in terms of activity in the labs afterwards.

Six of the students researched and taught themselves the technology using the relevant SAS Manuals.

Three students identified that using the SAS/AF and Frames technology would assist in achieving the highest mark levels. The most advanced student provided navigational and dynamic querying functionality that was roughly equivalent to some aspects of the SAS Visual Analytics product, despite apparently having not seen it.

Two students (who were comparatively advanced compared to their peers on the programme) used PROC SQL rather than the expected Base SAS Data step for selecting the data to be presented.

One student implemented his MIS using SAS Enterprise Guide.

The results for the programming and analytics part of the assessment were a 98% pass rate, with 77% of the students gaining an overall mark of 60% or higher. Taking into consideration the fact that the majority of the students had arguably chosen the BSc (Hons) Information Technology programme in order to avoid any programming, and had not shown any previous aptitude for coding, this can be considered a major success story in terms of both the levels of achievement, and the level of engagement and technical understanding demonstrated by the highly innovative solutions offered by some of the students.

These results suggest that further research into this “Learning by Exploration” approach is needed, and that it may be an approach that could bear fruit in other areas.

Bibliography


