

Gerald's Column *by Gerald Fitton*

Since my arrival on the Internet my correspondence has increased by a huge factor. The number of emails can be counted in hundreds per week and I am amazed at the way in which the publication of my email address in Archive generate revenue for the telecommunications industry. My web site continues to be visited regularly and I have received a short message from Demon which reports something called 'bandwidth usage'. They send these messages when the 'bandwidth usage' is above a certain level.

Referring now to BASIC Solutions in last month's Archive. From past experience I know that there are BASIC programmers who defend to the bitter end the systematic approach to a problem. In the past I have received correspondence from only a few who agree with the sentiments I express in that article about abandoning the totally systematic approach in favour of 'best guesses', intuition and a little thought. If you're on 'my side' then please let me know because I might begin to feel a bit lonely if I get too many complaints!

The defenders of ArgoNet continue to tell me that I am either biased or unfair. It seems that if you support ArgoNet then you do so with great gusto. It is good that ArgoNet has such staunch supporters – with support like that it won't be long before they have the best system on the net. To Argonet users I say, tell them how they can improve their system and I'm sure they'll get on with it.

Finally, the account of my years in employment have been received with such enthusiasm that I am encouraged to continue.

Summary

At University I amused myself by using the College computer to bet on the result of football matches and horse races as well as managing to use it to find the square root of 2.

In my first job I earned my living as an Engineer. I helped to make things like aeroplanes which worked – and I used their computer to play Solitaire.

In my next job I made profits, mainly for other people. Whilst there I used one of the first Sinclair computers (was it a Z80?) to gamble successfully on the metal futures market. I also used the same microscopic computer for stock control and cash flow analysis to the great benefit of the owner of the company. I was good at making profits for the company – but one morning (a few years after the old owner had sold out to an international plc) I woke up and decided that I didn't like the person I was becoming. A month later, after an argument with the new Chairman about the way we should operate, I was sacked.

The College

In my first job I made things; in my second job I made profits. By the time I came to the interview for my last job I wanted to make something different.

At the interview I expressed the view that I saw lecturing as an opportunity to make people. I think that, of the interview panel, only Bill, who was to be my Head of

Department, understood what I meant – and approved. Bill was into Education rather than mere qualifications. He was a people person and I like to believe that he saw in me things that I came to realise only when my students told me I had those qualities. There is no doubt in my mind that making people has been the most rewarding of my three careers. I have taken that opportunity to peddle wisdom and some have found the currency to buy it from me.

Computers at College

I started my job at Swindon College as a part time Lecturer in January 1983. This became a full time post in September 1983. The Archimedes was launched in 1988 and I had one of the first, an A305 which I later replaced with an A440. Now I have an A540 and an A4. I still don't have a RISC PC nor a StrongArm.

Between 1983 and 1988 I used a BBC B and later a Master to do much word processing and a few sums. From the beginning I decided to type up all my lecture notes at home. Soon I was producing my own handouts on a BBC B and an Epson FX80 8-pin dot matrix printer. The software I used was WordWise and, thanks to support from Paul Beverley's nascent company, I learnt how to use WordWise to good effect. I still have the books which Paul produced and I still marvel at the programs (now they would be called macros) which were available in the WordWise programming language.

I had limited funds and so I bought the printer years before I bought a disc drive; all my work was stored on cassette tapes – I still have the original tapes but I did copy the files to a 5½ inch disc when I got a Master!

In 1988 I bought an A305; later I exchanged it for an A440. Within months I was given a copy of the single tasking PipeDream 2 to review and I took to it immediately. When Acorn released RISC OS 2 the multi tasking version of PipeDream, PipeDream 3 was launched. The multi tasking Archimedes with its outline fonts and PipeDream 3 revolutionised the quality of my printed handouts. To take advantage of this new system I bought an expensive laser printer and became the envy of my colleagues who were beginning to use MS-DOS 3.

My students asked for extra copies of my handouts for their friends in other classes! To a large extent I owe my reputation for good handouts to the Archimedes; these handouts were my first experience of the 'people power' of students. I learnt a valuable lesson there.

Screening Tests

To the amazement of many who hear about it, Lecturers do things other than lecturing. One activity in which I became heavily involved was screening new students for what is called 'Numeracy' – but sometimes I wonder exactly what it was we were testing (see below). I want to tell you something about tests in a general way and in particular the tests which I devised and developed over the last ten years or so. I could not have developed these tests without PipeDream and the Archimedes.

Multi choice or Open questions

From the start I decided to go for multi choice questions rather than open questions. An open question is one where the candidate can write down any answer they like. An example of an open question is: "What is the result of adding 3 and 5?"; the candidate can write down any answer they can think of. I guess that you all know what multi choice questions look like but just in case you don't this is how it works. The question is the same, "What is the result of adding 3 and 5?" but the candidate is given a choice of answers such as 2, 6, 8, 10, 15, and has to tick the box adjacent to the answer they believe to be right.

I decided on multi choice questions for two reasons. The first is that the marking is totally objective. What I mean by "objective" is that anybody (almost regardless of their skill level) can mark the paper and all (accurate) markers will get the same number of correct answers. The second is that answers are either 'right' or 'wrong'. There is no need to look at the 'workings' nor to figure out the extent of the candidate's knowledge or ability from their part or incorrect answer. The third is a consequence of the second. A lot of papers can be marked quickly. Indeed I had pre printed pro forma answer sheets to give to the candidates. All that was necessary to mark the papers was a transparency with the correct answers highlighted. The transparency showing the correct answers was placed over the candidate's paper. I'm sure you'll see how easy it was to identify the wrong answers. Next year (well, it's all happened by now without me) the answer papers will be passed through the (one and only) College scanner and the right and wrong answers printed out and stored in the College computer.

Now, I know there are a lot of you who are in the education business who will be against multi choice questions for very sound reasons. Let me say that I think I know all of them but, if you want to 'have a go' at me then please read to the end of my article first.

Ranking Candidates

The first thing I did with the marked answer papers was to rank the candidates. I can almost hear the howls of protest from those who have been brain washed into believing that such a way of labelling candidates is both cruel and counter productive. Of course I disagree with you but, before you reach for your pen (or word processor), please read to the end of my article first. For the want of better words and in an attempt to avoid euphemisms I shall label those who got few right answers as 'weak' and those who got many right as 'strong'.

I entered all this data into my Archimedes computer. Indeed I entered the results first into PipeDream 3 and then later into PipeDream 4. Please note that I did not enter just the aggregate score but the way in which every candidate answered every question.

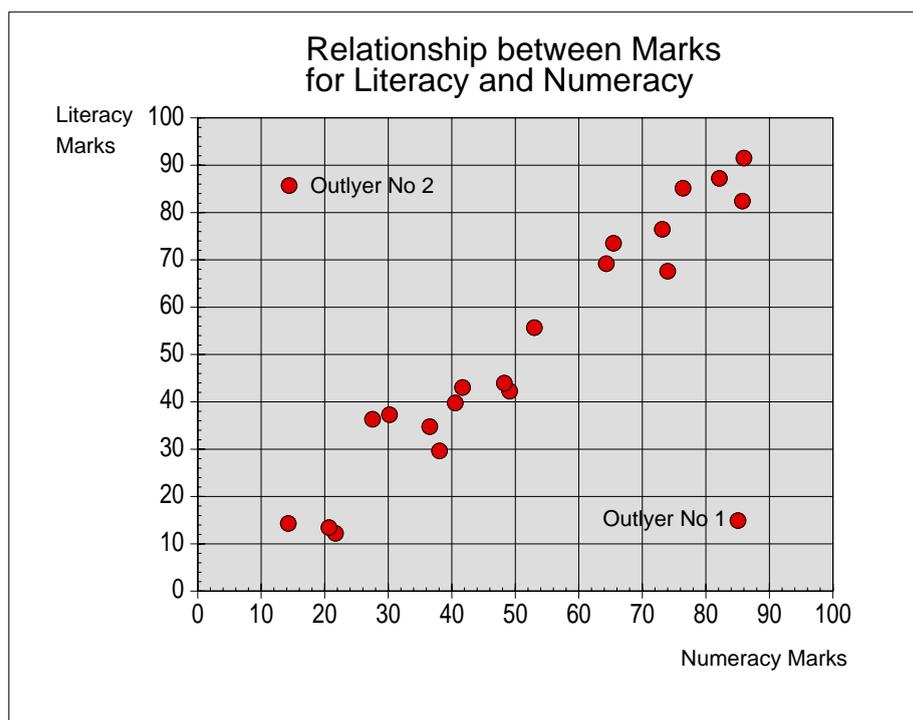
I shall come back to this shortly and then I shall tell you what we did with the results. Before I do so there is something else to tell you.

Numeracy and Literacy

I had such success with my 'Numeracy' screening tests in the first couple of years that the course team decided that they would run a 'Literacy' screening test as well. Persuaded by my arguments these were also multi choice and included comprehension of a written passage as well as the interpretation of advertisements and the subtle differences in the meanings of words. An example of the latter is to replace the word "nice" by one more appropriate from a multi choice selection.

Back to using the computer. Have a look at the drawfile below. For my own use I used PipeDream to create basic charts with little additional text but, for circulation when I produced my usual report, I added extra text before printing the result.

The results depicted here are not for any particular year and the drawfile shows only 22 points instead of the 100 to 150 candidates we tested each year. Initially, about eight to ten years ago, the marks in both Numeracy and Literacy were grouped in the middle range. More recently we found the results were much more polarised; candidates either did very well or very badly with few candidates getting near average marks in either Numeracy or Literacy. The scattergraph I have created shows the marks fairly evenly distributed between 10% and 90%; this corresponds to the situation about four or five years ago.



A couple of years ago we got a new Literacy Lecturer. She decided that the Literacy test we had been using for about six years was invalid and set her own. It required the candidates to write a short essay on "Why I came to Swindon College". The essays were marked by that Lecturer and the correlation with the Numeracy test was close to zero! I maintained that the marking was highly subjective but that was not my only criticism. We ran the old Literacy test a week later and the correlation came back to its old level. The correlation between the old Literacy test and the new test was near zero, in fact it was slightly negative. Of course I know the question you are asking. It is "What is it that our

tests are testing?" Let me avoid answering that just now.

A final couple of points about this graph. I have shown a couple of outliers which I've called Outlier No 1 and Outlier No2 (I'm never sure of the spelling of outlier but I use the one I do by consideration of 'outlying' as my root word). We always get some of these each year and there is always a good explanation.

Those of the Type 1 tend to have English as their second language and, at home they speak some other language. Those of Type 2 are even more interesting. Through additional diagnostic testing it transpires that these candidates have difficulty with many abstract concepts to different degrees. The worst cases have problems with the abstraction of colour (even though they are not colour blind) but generally the level of difficulty is that of describing the properties of such abstractions as kindness (is a certain action kind or unkind?) and justice (what is 'fair'?). Almost invariably they score well when they do an alternative numeracy test which includes no abstractions. $1 + 2$ is abstract. One apple plus two more apples is concrete.

Ranking the Questions

This is where my Archimedes (and PipeDream) really came into its own. Originally I applied the method which I describe below to study the teaching of A Level Maths students (in their second year) and I used old multi choice exam papers. It was so successful in that field that I extended the analysis to the screening tests.

I have said that I stored on my Archimedes in PipeDream not only the candidates' scores but their answer to each question. It is relatively easy to extract from that data a table similar to that below. You have to remember that the number of candidates would not be 5 as in the screenshot but between 100 and 150. Likewise, the number of questions was not 8 but about 70.

Ranking the questions

Question	1	2	3	4	5
Hard 8	-	-	-	-	Yes
7	-	-	-	-	Yes
6	-	-	-	Yes	Yes
5	Yes	Yes	-	Yes	-
4	-	-	Yes	Yes	Yes
3	-	-	Yes	Yes	Yes
2	-	Yes	Yes	Yes	Yes
Easy 1	Yes	Yes	Yes	Yes	Yes
	Weak				Strong

The Student

Let me explain to you what the table means. From left to right at the bottom of the table you will find the candidates in rank order with the weakest on the left and the strongest on the right. Probably you'll find that relatively easy to understand. What you might find more difficult is the concept of ranking the questions. In fact, how is it done? What is it

about the question which is measured and then ranked? The answer is that the candidates are used to rank the questions with the easy questions near the bottom of the table and the hard questions near to the top.

Question 1 has been answered correctly by all the candidates so it is an easy question. Question 8 has been answered correctly by only one candidate so it is the hardest question. In between we rank the questions according to how many candidates get the right answer.

Bad Questions

Why bother ranking the questions? Because it allows us to identify bad questions. In the jargon appropriate to this branch of Statistics bad questions are called ‘indiscriminating questions’ because they do not allow us to discriminate between weak and strong candidates.

There are two bad questions amongst the eight.

The obvious bad question is Question 1. It was answered correctly by everybody and so wasn’t worth including. It does not give us any information about the relative strength or weakness of our candidates. Even without the painstaking analysis which I carried out using PipeDream most Lecturers wouldn’t bother with Question 1 in future.

Not so easy to detect as a bad question is Question 5. It is bad because it doesn’t fit the pattern set by all the other questions. When testing for Numeracy a bad question might be one which asks how many people serve behind the bar in the Eastenders TV Soap Opera. Another example which was originally included was one asking candidates to estimate the length of a (picture of a) nail by using a pair of Imperial and metric scales which were printed on the question paper. The choices included 1 inch, 1¼ inches, 25 mm and 30 mm. The correct answer (called the ‘key’) was 1¼ inches and the others (called ‘distractors’) were classed as wrong; we got random results between the key and the 30 mm distractor. I was able to discover this because I stored all the candidates’ answers on the Archimedes and not just ‘right’ or ‘wrong’.

Identifying those who can answer Question 5 correctly might be interesting but that is not the point. The point is that the analysis shows that Question 5 does not test the same thing as questions such as 34 – 18. When we label Question 5 a bad question we are rejecting it because it doesn’t seem to be testing the same thing that the other seven questions are testing – whatever that is!

In the table I have made it quite obvious that Question 5 is a bad question. The two weakest candidates have answered it correctly and the strongest candidate has the wrong answer. Usually bad questions can not be spotted that easily. Anyway, there are degrees of ‘badness’ – the degree to which a question fails to discriminate is quantitative and not qualitative.

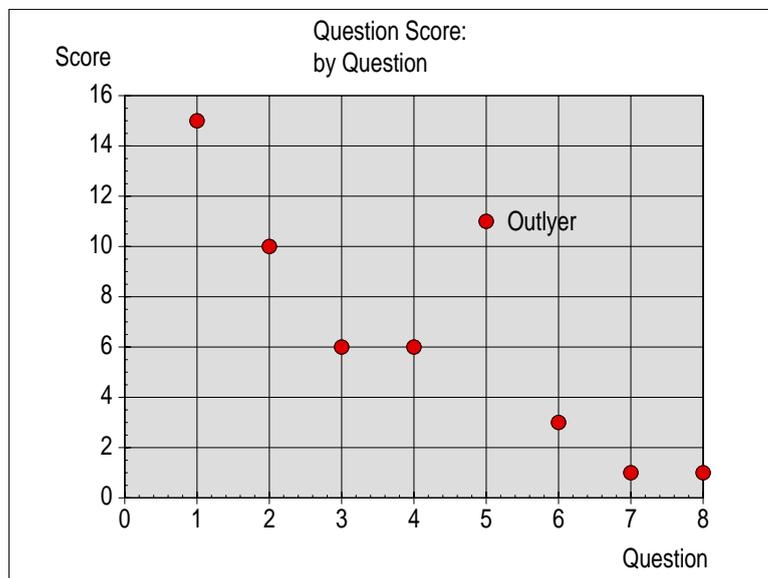
The table below shows one way of spotting bad questions. It is not quite the method I used but it is similar in principle. In addition what I did was to analyse the ‘distractors’ in an attempt to make the distractors equally distracting.

Ranking the questions						Score	
Question							
Hard	8	---	---	---	---	Yes	1
	7	---	---	---	---	Yes	1
	6	---	---	---	Yes	Yes	3
	5	Yes	Yes	---	Yes	---	11
	4	---	---	Yes	Yes	Yes	6
	3	---	---	Yes	Yes	Yes	6
	2	---	Yes	Yes	Yes	Yes	10
Easy	1	Yes	Yes	Yes	Yes	Yes	15
		1	2	3	4	5	
		Weak		The Student		Strong	

Questions answered correctly by the weakest candidate, Candidate 1, are given 5 points. Those answered correctly by Candidate 5, the strongest, are given 1 point. The points scored by each question are totalled. These values are shown in the last column of the table above under the heading "Score".

Graphical Analysis

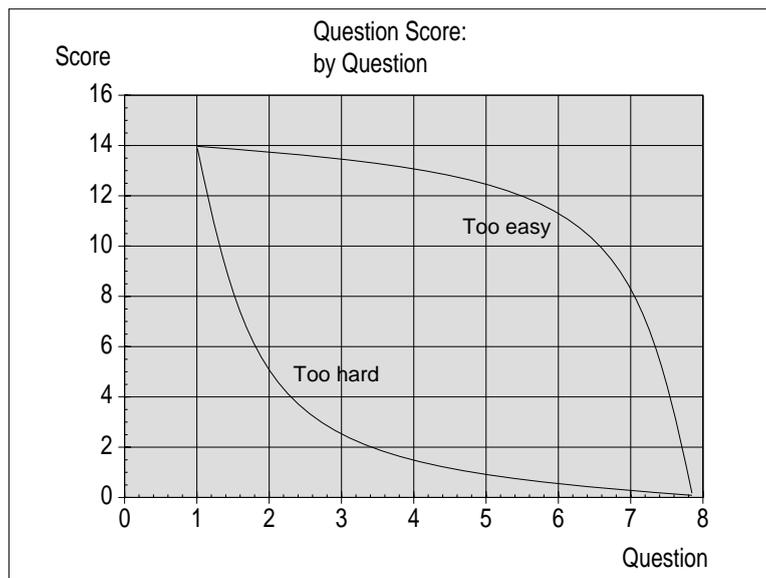
The graph shown below is a plot of the question score against the ranked question number. The general pattern of the graph is that it is close to a line from top left to bottom right. Question 5 stands out from the general trend of the graph as an outlier. I used a graphical method similar to but not identical with this one to identify bad questions.



To Many Hard/Easy Questions

One of the problems of writing a test paper is that you mustn't include too many easy questions nor too many hard ones. Ideally you need a set of questions which will discriminate well at all ability levels. The line graphs shown below demonstrate the sort of

Score/Question graph which you get if you have too many easy or too many hard questions.



If you have too many hard questions then you may be able to identify the subtle distinctions between the strong and strongest candidates to a high degree of accuracy but you will not have much data which relates to the weakest candidates. To include too many easy questions is an equally bad mistake. The ideal Question/Score graph (for our purpose) is one which is a fairly straight line from upper left to lower right.

The technique I actually used to fine tune the Numeracy test paper was not exactly as described above but the fundamentals of the method are as I indicated. What I did in addition was to fine tune the distractors.

All the analysis on both the Numeracy and the Literacy questions was done by me at home using an A440 and PipeDream. After a while I was able to discern general principles for the writing of good questions (and useful distractors) and advise the Literacy team what principles to apply to their questions.

What were we testing?

I know what the Numeracy paper contained and, after a year or so, I had a set of questions which produced consistent results. There were no bad questions – though some were more reliable discriminators than others.

The Numeracy papers contained mainly abstract simple arithmetic. The paper started with the addition of one and then two digit integers, working up through subtraction without the need to ‘borrow’, to three digit addition with carry and subtraction with borrowing. In the next section there were questions which involved proportion, ratio and percentages. Following that values had to be read off simple graphs. Another section involved approximations. In this section the multi choice answers never included the exact answer to the sum so the candidate had to find the nearest answer. Finally we included a section of concrete applications which included distance, area, volume, money, and compound

measures such as speed (m per second) and prices (50p per kg). We did not test numbers larger than 999, nor negative numbers nor fractions other than $\frac{1}{4}$, $\frac{1}{2}$ and $\frac{3}{4}$.

We found that questions including pi (area or circumference of circles) were unreliable indicators, fractions other than those mentioned were unreliable indicators, sums involving more than four digits were unreliable as were many other things I won't bore you with. By unreliable I mean that the Score/Question graph showed many outliers. I put this down to the emphasis given to teaching these things at the various schools in our area. Some taught the uses of pi and others didn't. What I wanted to test was whether the candidates had been able to learn, remember and use the things they had definitely been taught. If they hadn't been taught about pi then they wouldn't be able to answer a question which needed it – however 'clever' they were.

So what were we testing? One thing which nagged at the back of my mind for a long time was the high correlation with the Literacy test and my insistence that the Literacy test should be tailored so that it yielded results which correlated well with the Numeracy test. Did I change the Literacy test to prove my hypothesis? Had I been unfair to the new Lecturer?

I would like your opinions before I tell you mine – I've already dropped you one hint a couple of paragraphs back. Write to me or send me an email with your views and, in a month or so, I'll let you know mine.

What did we do with the results?

One thing we did was to offer extra lessons to those with low scores. The take up of the extra lessons was about 50% and those who took them up improved rapidly under the guidance of sympathetic teachers with low class sizes. In the second half of the second year I ran what might be called a 'progress' or 'exit' test containing similar questions. Those who had attended the extra lessons improved their score considerably; many were obtaining Distinctions in their main core subjects. It is my opinion that these students realised how hard work could improve their performance and were highly motivated to do as well as they were able.

The High Scorers

About six years ago I discovered something else from the screening test results. It came as such a shock to me that I didn't believe it was due to anything but chance for another two years. I looked at the screening test scores and identified the top 15% of students (about 15 to 25 students) and checked on their achievement with us after the two years they'd been with us. To my utter astonishment most of the students never finished the course. Two students finishing was the highest we ever achieved. It was more usually zero. Furthermore, whilst those students were with us they were all sorts of trouble ranging from skipping lessons, handing in work late or not at all, to being disruptive in class.

During the last six years I spent a couple of years disbelieving the evidence, then I spent two years trying to discover what was going wrong. I think I might have got somewhere because all but one of the high scoring students who completed were members of my tutor group and had massive trouble with the Lecturers.

During the last two years I tried to get the Educational Psychologist interested. His (free of charge) contribution was that it was all down to Student Motivation. Unfortunately the College wouldn't pay him (he was a Part Time consultant) to investigate my findings in greater detail so I never got a real expert to look into it.

You will appreciate that once students have left it is difficult to find out why they left if only because they aren't around any more to ask them. What I did was to identify these students from the screening test and then, well before they left, I ask them what they think about the course, etc. One common complaint was that their (often original) work was marked down for petty reasons "just because it isn't what the Lecturer said". Another was that they were "Bored". Like many words, the meaning of "Bored" has changed with time. When a student says "Bored" it doesn't exactly mean that they find what they are doing uninteresting. The modern usage by students and maybe other young people of "Bored" implies a lack of understanding of what is important to the student by those in authority – in management jargon there is a 'communications gap'. I would appreciate your views particularly if you have experienced this phenomenon.

I have recommended that we merge the Numeracy and Literacy tests mixing the questions and that we expand it to include behavioural and motivational questions and other non academic questions such as sporting or leisure interests (to hide the true reason for the test) and then analyse the data in different ways for different purposes by selecting only those questions relevant to the person doing the analysis. This won't be taken up because the Lecturers don't have the motivation to do anything more about it.

Finally

As part of my preparation for retirement I decided to undergo psychological testing. Paul has referred to my modesty and will not be surprised to learn that the psychologist concluded that I am not motivated unduly by money, sex, security, etc, but Paul (and you) may be surprised to learn that for most of my life I have been motivated by an irresistible craving for power, power to control both the events and the people in my life!

With a great deal of hindsight I have to agree that the psychologist is right. I have never been a good team player unless I have been the leader. My classroom and my tutorial groups have always been run as an autocracy with me as the autocrat. When I took up my third career at College I resolved that I would never seek promotion out of the classroom. It was my intention to be a good teacher and not get involved in the management hierarchy. I didn't want that power – "Been there, done that, got the T shirt, read the book, seen the film, didn't like it!" I exercised my thirst for power in the classroom.

There is a saying that "all power corrupts". In my opinion it is only to the extent that the power is not absolute that it corrupts. Hence it is my view that absolute power does not "corrupt absolutely" as in the saying, but that absolute power, such as that wielded by God, liberates the holder to become as charitable (see a previous Archive) as they wish to be.

I hope that in my classroom I have been a benevolent dictator – I have certainly out performed all my colleagues both as a teacher and as a tutor to a statistically significant extent, in some specific measurable management statistics (eg retention rates) by a factor of two. As every good teacher knows, if you can get your relationship with the student right then, for the student, everything which you believe is possible for them (the student)

is achievable by them.

After getting to know me, most of my students have liked me sufficiently to believe in my judgement of their potential – and have achieved much more than they, or my colleagues, had believed possible. Faith by students in their Tutor and Lecturers can move mountains of self depreciating prejudice. The Faith I had in their potential was my freely given gift to them. Faith (in anything) is one of those (three) things which “abideth”. Once you have the real thing then, unlike Coke, it lasts forever.

Quo Vadis?

So, is my retirement the end of my struggle for power over my colleagues and charges? Will I have to give in gracefully and accept a less powerful role as I live out the remainder of my life? The psychologist believes that I shall have a real problem and that it will grow more acute as time goes by. He gave me no help with its solution so your suggestions will be most welcome.

Address

Snail mail, fax, email and web site are that of Abacus Training listed on the back cover of Archive. You can call me “Gerald”!