An investigation into students' global A-level Mathematics performance and their performance in routine and non-routine questions

A project supported by the

Departmental Investment and Restructuring Fund, University of Leeds

John Monaghan

School of Education

University of Leeds

Acknowledgements

This project was a joint venture by staff at the University of Leeds and the University of Plymouth. University of Leeds staff were supported by the Departmental Investment and Restructuring Fund. The project staff were:

University of Leeds, UK	University of Plymouth, UK
Peter Johnson (research assistant)	John Berry
John Monaghan	Wendy Maull (research assistant)

The project team would like to thank the Research & Evaluation Division of the University of Cambridge Local Examinations Syndicate for providing A-level Mathematics scripts.

Several schools in the Plymouth and Leeds regions assisted the project team. Special thanks is given to Susan Metcalfe of Heckmondwike Grammar School.

NB This report is written in a style that allows non-UK educationalists to appreciate details, e.g. A-level grade details.

An investigation into students' global A-level Mathematics performance and their performance in routine and non-routine questions

Abstract *This study concerns student performance in pre-university examination questions. In particular whether lower attaining students in mathematics examinations generally gain their marks on routine parts of questions? This is an important issue because routine questions could be awarded fewer marks if algebraic calculators are allowed in examinations. Students' scripts in a recent mathematics examination were examined in an attempt to evaluate this question. The results are not conclusive but indicate that a problem of this type does exist, though the nature and location of the problem is not as straightforward as expected.*

Introduction

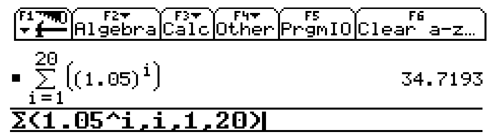
Our starting point is the question: do A-level Mathematics[1] students who attain lower pass grades (D and E) generally obtain these grades by answering 'routine' parts of A-level Mathematics questions? Routine questions may be viewed as those for which students may be expected to execute a rehearsed procedure consisting of a limited number of steps. Problems in characterizing routine questions are considered later. The next three paragraphs explain the rationale for and import of the study.

During the period 1994-1996 the then Schools Curriculum and Assessment Authority (SCAA) set up a number of working groups investigating possible consequences of student use of a new generation of algebraic calculators on A-level Mathematics questions and papers. One important debate was whether such use would accentuate the difference between higher and lower attaining students, e.g. between those attaining A-level grades A & B and those attaining grades D & E. An example should clarify matters.

A typical question on geometric series may start by a request to evaluate $\sum_{i=i}^{20} 1.05^{i}$

[1] Advanced level (A-level) Mathematics is the most common senior public examination for students in the UK. It covers considerable algebra and calculus of a single variable. There are five pass grades, A to E. Examinations are set and marked by independent institutions called Examination Boards. Examination sheets are called papers.

and then proceed to a question on compound interest, e.g. “If I invest £550 at a rate of 5% per annum, how many years must I wait until I have more than £1000 in this account?” It should be noted that the new generation of algebraic calculators can perform the first part of this question, e.g. the TI-92[2] screendump to the right. In the ensuing discussion everyone initially assumed, as a generality with exceptions, that students attaining lower A-level grades learnt how to do the first routine part but would have difficulty with the second non-routine part. At the next meeting, however, the discussion continued with several people saying they were not sure that this really was the case. There appears to be no literature of direct relevance in this area..



Now if lower attaining students generally obtain the majority of their marks on routine questions and if, as seems likely, such questions are allocated a relatively smaller share of the total mark scheme when algebraic calculators are permitted in examinations (see Monaghan (to appear) for a discussion of this issue), then these students will find it more difficult to pass these examinations. Mathematics is already considered a difficult subject at senior school level (see Fitz-Gibbon & Vincent (1994, p.23) for UK data) and we would be extremely concerned if mathematics examinations became more difficult to pass.

Methodology

To address the question we analysed the performance of students with different A-level grades (A, B, C, D & E) in questions which have routine and non-routine parts. An A-level Examination Board provided us with the scripts from a recently marked Pure Mathematics paper. A Pure Mathematics paper was chosen, rather than a Mechanics, Statistics or Discrete Mathematics paper, since Pure Mathematics is the core for all mathematics options and because it is the area most likely to be affected by algebraic calculators (see Monaghan (to appear) for a discussion of this). Over 300 scripts from an almost equal number of male and female students who obtained scores at the boundaries of the A, B, C, D and E grades were provided.

Each question part was coded as routine or non-routine (further details below) and the marks for the question parts were adjusted in accordance with our expectations of what

[2] Texas Instruments’ TI-92 calculator is an example of an algebraic calculator.

future marks, where routine questions were allocated a relatively smaller share of the total mark scheme, would be like. Students' scripts were then remarked.

The A-level paper and mark scheme(s)

The paper used was the first of six papers. It was one of three papers that all students following a popular modular A-level scheme had to take. There were four questions each worth 15 marks. Questions 1, 2 and 4 had four parts. Question 3 had five parts. We reproduce question 2 below as an example of the kind of question asked. The original marks are given in square brackets.

The gradient of a curve is given by $\frac{dy}{dx} = 3x^2 - 8x + 5$. The curve passes through the point $(0, 3)$.		
(i)	Find the equation of the curve.	[4]
(ii)	Find the coordinates of the two stationary points on the curve. State, with a reason, the nature of each stationary point.	[6]
(iii)	State the range of values of *k* for which the curve has three distinct intersections with the line $y=k$.	[2]
(iv)	State the range of values of *x* for which the curve has a negative gradient. Find the *x*-coordinate of the point within this range where the curve is steepest.	[3]

The other questions concerned: Q1, trigonometry in context; Q3, coordinate geometry (lines, circles and ellipses); Q4, integration in context (comparison of exact and numeric methods). We classified each question part as routine (R) or non-routine (N) and obtained: Q1 (R, N, N, N); Q2 (R, R, N, N); Q3 (R, R, N, N, N); Q4 (R, R, R, N). The division of marks for routine and non-routine parts was 30 marks each. Various alternative mark schemes were developed, all adjusting the mark ratio so that routine parts of questions scored fewer marks. The agreed version left each question with 15 marks, left the mark allocation of Q1 unchanged but adjusted the others so that routine parts totalled 23 marks and non-routine parts totalled 37 marks. The parts of Q2, for example, were allocated 3, 4, 4 and 4 marks respectively.

The grade boundaries for this paper were: A, 40; B, 33; C, 27; D, 21; and E, 15. In keeping with A-level conventions A, B and E grade boundaries were determined by examiners' judgement while C and D boundaries were fixed at equal intervals between B and E grade boundaries.

Results

311 scripts (63, 63, 62, 62 and 61 at grades A, B, C, D and E respectively) were remarked to the new mark scheme. The use of statistics in this study must be carefully examined for much of the data is far from independent (consider, for example, the relationship between the total on the original mark scheme and the total on the revised mark scheme). The statistics which follow are intended to give the reader a feel for the general patterns in the data. Three aspects are examined here: the overall scores; the proportion of marks to routine and non-routine parts of questions; and factor analytic results suggesting that students follow through whole questions.

The scatter diagram, figure 1, of old and new totals illustrates the general pattern. The ranges overlap but the ranges from the original grades retain their hierarchical structure. The new totals are generally lower than the original totals. In fact of the 311 student scripts examined 297 obtained lower scores from the new mark scheme, 11 scores remained the same (3, 3, 2 and 3 from A, C, D and E grade students respectively) and three obtained higher marks (a B and a D grade student obtaining one more mark and a C grade student gaining three more marks). This indicates that increased emphasis on non-routine questions leads to a general lowering of the overall marks obtained.

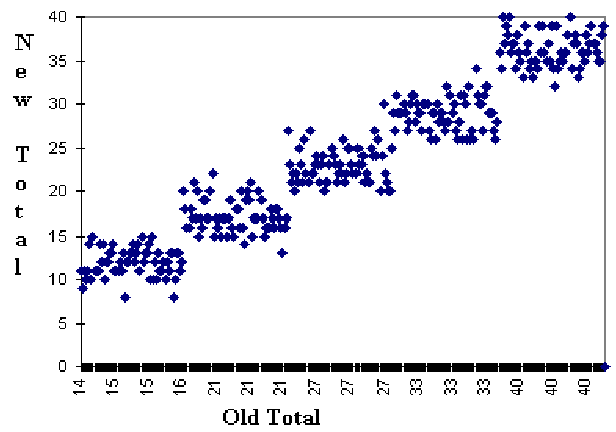


Figure 1, scatter diagram showing total scores under both mark schemes

Figure 2 displays the mean marks obtained by groups of students at each grade level for:

- Old routine (O-R) parts (out of 30)
- Old non-routine (O-N) parts(out of 30)
- New (revised) routine (N-R) parts (out of 23)
- New non-routine (N-N) parts (out of 37)

	O-R	O-N	N-R	N-N
A	26.3	13.7	19.9	16.4
B	23.5	9.5	17.4	11.2
C	19.9	7.1	14.8	8.1
D	15.8	5.2	11.1	6.2
E	11.8	3.2	8.3	3.6

Figure 2, mean marks over grades

Note that each column decreases with decreasing grades (hardly surprising) and that mean marks obtained for routine parts of questions are consistently greater than mean marks obtained for non-routine parts of questions even though there were more marks for non-routine parts of questions in the revised mark scheme. This may be interpreted as evidence that students attaining at all pass grades gain more marks on routine parts of questions An examination of the ratios O-R:O-N and N-R:N-N is interesting. For grades A-E we get, respectively: 1.9, 2.5, 2.8, 3.0, 3.7 and 1.2, 1.6, 1.8, 1.8, 2.3. The decrease in the second list, relative to the first list, clearly mirrors the higher weighting given to non-routine parts of questions in the revised mark scheme. The increase in both lists, however, may be interpreted as evidence that higher (respectively lower) attaining students gain proportionally (to their overall mark) more (respectively less) marks on non-routine parts of questions.

Principal component analysis of both the old and the new scores yielded 6 factors with eigenvalues greater than one. In both old and new scores the question parts which loaded significantly on the factors were as follows: all parts of Q2; all parts of Q4; Q1 parts i, ii, and iii; Q3 parts i, ii, iii, and iv; Q3 parts ii, iv and v; Q1iv, Question 1i (negatively) and Q2i. The first five factors suggest an interpretation that the correlations of the scores of the parts within a question dominate the analysis, i.e. if you do well on one part of Q2, you tend to do well on all of it.

Grades and routine questions

So, do lower attaining students gain a substantial proportion of their marks on routine parts of questions, compared to higher attaining students? The results are not conclusive but they are not without interest. Before considering them we address an aspect of their surface validity. The results arise from at least two semi-arbitrary decisions: the categorization of parts of questions as routine or not (and the dichotomy implicit in this categorization); the weightings given in the revised mark scheme. These are important factors to bear in mind but it should be noted that both decisions were made after considerable debate by a group of people with considerable experience of the type of examination paper in question.

The overall lower scores obtained from the revised mark scheme and the distribution of mean marks over routine and non-routine question in both mark schemes clearly indicate that all students score substantially more marks on what we have designated routine parts of questions. The increasing ratios of routine to non-routine mean marks over grades in both mark schemes, however, does provide evidence that lower attaining students do obtain proportionally more of their marks on routine parts of questions. Looking at the scatter diagram in figure 1, however, it would not appear that this would make a substantial difference to the overall grades (if they were still determined by the same judgement/equal interval rubric) given the small ’new total’ overlap over old grades. Indeed, several experiments at reallocating grades under the revised mark scheme were conducted and none of these resulted in more than a 10% reallocation of grades. The outcomes are not recorded in this paper for fear of introducing another semi-arbitrary element.

The principal component analysis results alerted us to a possibility we had not, but should have, anticipated: that many students, at all levels of attainment, exhibit a propensity to follow a question through. This may have many bi-causal connections with other influences, including a familiarity with a specific content area – indeed a dialectic may exist between attainment and familiarity with a range of content areas. Again we must view these results with caution, due to regularity conditions implicit in the analysis and a lack of prior hypotheses, but five out of six located factors indicating such a propensity in students certainly deserves consideration.

What this and other results lead to is the need for further investigations. Two such investigations are a refinement of what is meant by the term ’routine question’ and a more realistic interpretation of curriculum development vis à vis assessment development. We turn to a considerations of these now.

Problems in characterizing routine questions

A problem with this study is that it implies that routineness is located in a question rather than being a psychological construct of the relation between an individual, or group, and a question. The latter appears more valid. Indeed, the psychological relation is likely to be a socio-psychological relation. Routineness as located in a question is, however, one way to approach this study. This study's origins grew from a SCAA working party examining examination questions and we had the opportunity to explore this issue by examining completed A-level scripts without exploring classroom situations. Rather than viewing our approach as flawed we choose to view it as exploring one avenue of routineness.

There are very few readily available references to routine questions in the mathematics education literature. An important paper, which addresses a similar level/type of mathematics to this study, is Seldon, Mason & Seldon (1989). They make a distinction between problems and exercises and view, in a similar way to the above socio-psychological relation, a problem having two components: task and solver(s). They view 'cognitively non-trivial' problems as those where "the solver does not begin knowing a method of solution". and note that traditional calculus courses contain few cognitively nontrivial problems. They note that 'tasks' require skills and are divided into parts, algorithms, sample solutions and examples and that many problems are made routine in this way. This accords with our own experiences of UK A-level mathematics classes. Seldon et al, however, do not define what they mean by 'routine'.

Boaler (1997) explored a range of issues from two schools with strikingly contrasting ethoses and teaching methods. We focus here on her analyses of pre-A-level students' performance on conceptual/procedural questions, which may be viewed as a form of the non-routine/routine division. She defines procedural questions as "those questions that could be answered by a simplistic rehearsal of a rule, method or formula." Conceptual questions were viewed, in contraposition, as questions which require "the use of some thought and rules or methods committed to memory in lessons would not be of great help". Boaler claims that conceptual questions are more difficult, for students, than procedural questions, and descriptive statistics support this view. Notwithstanding the fact that cursory summaries do not do justice to Boaler's work it can be said that performance on conceptual and procedural questions shows that similar overall results in examination performance may be obtained in different ways (ratios of success in the two types of questions) under different school ethoses and subject teaching methods.

Boaler's work raises the obvious question of the relationship between routineness and teaching methods. Nagy et al (1991) examine the relationship between test content and instructional content at the level of High School calculus. It would be in injustice to characterize their study as purely quantitative as their foci are intent of instruction, nature of materials and operations. However, their analysis of assessment activities on a six-category system, ranging from skills to situational problems, showed wide variation in teachers' emphasis, especially at the skill level. This calls attention to the importance of further studies on instruction to which our study cannot make a contribution.

A parallel question to the 'routine/non-routine' distinction is "What makes one exam question more difficult than another?" (Fisher-Hock et al. (1997)). This is the starting point of the Question Difficulty Project (QDP) which examined UK mathematics and other subject examinations taken by 16 year old students. The project examined a model of question answering based on reading, application and communication through protocol analysis. The study notes the difficulty of both social and mathematical language, to the presentation of answers and the concomitant recording of steps (the latter two being particularly important for future work with algebraic calculators). Trials of variations in mathematics questions revealed 22 sources of difficulty, from command words to irrelevant information. The import of this for our study is the sheer number of factors impinging on what might make a question difficult (routine) or not.

This study and the work of Boaler alert us to the issue of context. Context is a term in mathematics education that is particularly difficult to define. Indeed, we believe that no definitive definition can be produced. It was the original intention of our study to examine contextual/non-contextual questions as well as routine/non-routine questions but the problems of finding real context questions and the problems of characterizing context questions forced us to focus on the more manageable routine/non-routine distinction. We leave the question of context to further research.

Three further thoughts on routineness:

The relation between experience and the examination question is metonymic, at the level of syntax rather than meaning[3], i.e. the form of words used in a question evokes a particular response. Arguably the most routine question at this level in the UK is, *find the equation of the tangent to __ at the point __.*

Students opinions of what are routine and non-routine questions are important, though it must not be assumed that students share a collective meaning of the term. We asked 100 students who had studied the paper we examined, in the course of their revision, to categorize the question parts, see figure 3. These responses show: only partial agreement with our categorization; several parts are neither generally perceived of as either routine or non-routine; greater apparent accord on routineness than on non-routineness.

	1i	**1ii**	**1iii**	**1iv**	**2i**	**2ii**	**2iii**	**2iv**	**3i**	**3ii**	**3iii**	**3iv**	**3v**	**4i**	**4ii**	**4iii**	**4iv**
total ’routine’	**98**	**81**	**47**	**4**	**99**	**97**	**32**	**40**	**95**	**87**	**71**	**47**	**49**	**93**	**69**	**87**	**70**
total 'non-routine’	**2**	**19**	**52**	**94**	**1**	**3**	**68**	**59**	**5**	**13**	**29**	**53**	**51**	**7**	**31**	**13**	**30**
Our categorization	**R**	**N**	**N**	**N**	**R**	**R**	**N**	**N**	**R**	**R**	**N**	**N**	**N**	**R**	**R**	**R**	**N**

Figure 3, student classification of routine and non-routine parts of questions

Finally we discussed routineness with the relevant mathematics officer of the Examination Board whose paper we used. They had a policy for question design: “In a standard question of four parts, the first two parts require an application of a standard algorithm, the next two parts require insight.” This may be construed as a form of the routine/non-routine division and accords quite closely with our categorization in the paper considered.

Curriculum development and assessment development

This study is located in what may be termed ‘traditional’ examinations. Common themes of contemporary research in assessment are the purpose and validity of assessment and alternatives to traditional forms of assessment (see Niss, 1993). We applaud this but

[3] Suggested by T Rowland in personal correspondence on the nature of routine questions.

traditional forms will continue to persist in many countries and investigations into how students perform in them are important. This will be the case with the UK A-level examination (see http://www.qca.org.uk/aframe.htm and go to *A/AS subject core consultation*, then to *Mathematics*).

In the medium and long term scenarios, when algebraic calculators become commonplace in developed countries' classrooms, it is possible that the curriculum will develop to incorporate the potential offered by these calculators (see Browne & Ellis, 1997). Ideally examination questions, if not the form of examinations, will change in line with curriculum developments. In this projected setting the current study is partially misplaced as an attempt to second guess the future without taking account of parallel curriculum and assessment development. It is difficult, however, to envisage how a realistic 'future' experiment might be conducted for traditional examinations.

In regions that permit less rigid senior school examinations the potential problems exhibited by UK examinations are not necessarily present. Where teacher-generated examinations are allowed there is scope for simultaneous curriculum and assessment development. In some schools in Austria, for example, where regular use of algebraic calculators is made teacher-generated 'parallel' calculator and non-calculator examinations take place (Browne & Ellis, 1997).

Conclusion

There are many issues in this study which require further investigation: the effect of teaching and learning styles and of instructional content on students' performance in questions and question parts; an analysis of what makes an A-level question difficult regardless of apriori categorizations such as routineness; an exploration of the factors involved when students follow a question, of several parts, through.

Regardless of further research there is a need, in countries with traditional senior school examinations, to take the indicators in this study seriously lest, when algebraic calculators are allowed in examinations, that the examinations become more difficult to pass for a class of students. Although this study in not conclusive the indicators are such that we cannot ignore the equity issue present.

A suggestion for further study

The inherent equity issue alone warrants a further exploration of issues raised in this report. Because the equity issue concerns advanced calculators further studies should explicitly consider calculator issues. The discussion on characterizing routine questions suggests a number of possible foci for an extension to this study:

(1) a simultaneous focus on task and solver

(2) a focus on teaching and learning styles

(3) a focus on instructional content and test content

(4) what makes an A-level question difficult

(2) and (3) are of considerable potential importance but are likely to be labour intensive in terms of classroom observation. For practical reasons it is suggested that (1) and (4) are explored. The suggestions below are influenced by the Question Difficulty Project. By focusing on difficulties students have with questions it both skirts many of the problems raised in trying to determine what routine questions are and gives a possible way to compare difficulty issues concerning calculator and non-calculator use. The following is a mere sketch of a possible research study.

Objective: to locate factors which make an A-level question[4] difficult with and without an advanced calculator.

To address this students must be observed doing mathematics. Concurrent protocols, Ericsson & Simon (1993), would appear to be more useful than retrospective protocols to locate the temporal place of student difficulties in the sequence of actions involved in answering a question.

Questions used should be neither too difficult or too easy for the students to answer since questions at either extreme will not locate sources of difficulty. A relevant *facility level range* should be investigated. Questions should be real examination questions because examinations have evolved with a specific style and accompanying language and stylistic and language issues are likely to affect question difficulty. Relevant questions to be piloted to discover their facility level with a representative sample of students. In order to

[4] Or, say, an Advanced Placement Calculus (USA) question

assess the role of advanced calculators in students' answers the questions should also be examined to ascertain whether advanced calculators are likely to assist students in answering questions or not. At this stage in the conception of the project it seems desirable to include questions for which advanced calculator usage aids students as well as questions for which advanced calculator usage does not aid students in answering questions.

Initial design to be informed by a model which can be applied to student question answering. The QDP's model to be piloted and evaluated as a suitable analytic tool. A simplified version of this without feedback loops is shown below.

Reading the question	Recognize	Understand	Plan
Application	Extract	Execute	
Communication	Record	More steps?	Check
Finish			

Concurrent protocol analysis is time consuming. The number of students involved is thus likely to be small, the exact number dependent on future discussions and the number of partners in this project. At this stage in the conception of the project three factors appear to be important in selecting students, classes and schools/colleges: representativeness; student attainment levels; use of advanced calculators. These are now considered .

Given that the sample of students would be small, they should not be put forward as representative of a student population. However, obvious atypical students, e.g. gifted, should not be included in the sample. Similar consideration should apply to schools/ colleges involved. Following up the current project, however, it would appear important for the sample to include students whose expected grade ranged over the A-E pass grades. Regarding the use, or not, of advanced calculators and student attainment levels it would appear important for the sample to include students who were adept with advanced calculators as well as those who did not use them and for the sample to include students from the full range of expected pass grades.

As this possible extension project is viewed as a collaborative project with partners as yet unknown further details are left for discussion.

References

Boaler, J. (1997) *Experiencing School Mathematics*. Buckingham: Open University Press.

Browne, R. & Ellis, W. (1997) Type, purpose and validity of assessment, pp. 63-70 in J. Berry, M. Kronfellner, B. Kutzler & J. Monaghan (eds.) *The State of Computer Algebra in Mathematics Education*. Bromley: Chartwell-Bratt.

Ericsson, K. A. & Simon, H. A. (1993) *Protocol Analysis: Verbal reports as data*. Oxford: Oxford University Press.

Fisher-Hoch, H. & Hughes, S. (1997) What Makes Mathematics Exam Questions Difficult? http://www.leeds.ac.uk/educol/BEID.html, then click on *advanced query*, then type 000000338 in the *anywhere* box.

Monaghan, J. (to appear) Some issues surrounding the use of algebraic calculators in traditional examinations. *International Journal for Mathematics Education in Science and Technology*.

Nagy, P., Traub, R., MacRury, K. & Roslyn, K. (1991) High School Calculus: Comparing the Content of Assignments and Tests, *Jnl for Research in Math Education*, **22**(1), 69-75.

Niss, M. (1993) Assessment in Mathematics Educations and its Effects: An Introduction, pp. 1-30 in M. Niss (ed.) *Investigations into Assessment in Mathematics Education*. Dodrecht: Kluwer Academic.

Selden, J., Mason, A. & Selden, A. (1989) Can Average Calculus Students Solve Nonroutine Problems? *Journal of Mathematical Behaviour*, **8**, 45-50.