

Paper for AEA-Europe Conference, Hisar, Bulgaria, November 6-8 2008

**EXAMINATION PILOTS:
WHERE CAN THEY TAKE US?**

By
Roger Murphy and Andrew Noyes
School of Education,
University of Nottingham

Examination Pilots: Where can they take us?

1 Introduction

Tests and examinations play a crucial role in shaping education systems in many countries throughout the world. Their significance goes well beyond the results that are derived from them as they also tend to exert a strong influence on the educational experiences, ambitions, and motivation of students, parents/guardians/carers and teachers. For these and for many other reasons they are subject to frequent reviews, reforms and some experimentation. This paper will reflect upon one instance where a set of new examinations are being piloted prior to this possible extended use with students in England. We intend to use this example as the basis for reflecting upon the role, function and challenges of conducting future pilot studies of examinations.

The Cabinet Office of the UK Government recently conducted a review of the use of 'pilots' in informing policy-making (Government Chief Social Researcher's Office, 2003). This review highlighted a range of benefits and challenges in using pilots as a step towards making policy decisions. Although the term pilot is now used quite widely, it is sometimes used to describe quite different types of activities. It is helpful therefore to consider exactly what is meant by a pilot and in the case of 'examination pilots' consider what aspects of the examinations are being piloted, what questions pilots are intended to address, and the ways in which it is intended to use evidence from pilots to inform future decision making. Such a reflection will inevitably cause us to address the broader issue of the different ways in which the suitability of educational examinations can be evaluated. In the particular pilot we will be reflecting upon, the examination reform relates directly to concerns about student learning, the nature of the mathematics curriculum taught in secondary schools in England, and the ways in which individual students can progress in terms of their learning of mathematics at different stages of their education. Such questions, we argue, demand much broader evaluation criteria and methods than those used in evaluations of educational examinations, which often in the past have relied on narrow psychometric measures and assumptions. Common examination at 16+ pilots (that led in time to GCSE), for example, back in the 1970s, were evaluated largely on the basis of whether the examination papers produced normal distribution results curves rather than addressing more fundamental questions about their suitability for comprehensive schools, teaching, and a wide range of learners.

First then we will introduce our 'examination pilot' case study and then we will return to some broader discussion about the role of 'examination pilots' in helping to inform critical decisions about examinations reforms.

2 Some background to the Mathematics Pathways Pilots

The 'examination pilots', which our research study is evaluating have emerged from a complex set of debates and circumstances. Since the late 1980s participation in A level mathematics in England had been in decline (although there are signs that this trend is slowing and even reversing). Following a major review of qualifications in 1996 led by Sir Ron Dearing new A Level qualifications were introduced as Curriculum 2000. This change was widely seen as a 'disaster' for mathematics as a large number of students failed their AS, the first half of their A level award and so dropped out. This apparent crisis in mathematics precipitated renewed concern over mathematics (and science) qualifications and considerable discussion and advisory activity resulted. A focused review of 14-19 mathematics (Smith, 2004) took place alongside a broader review of 14-19 qualifications by Tomlinson (2004). Tomlinson's proposed diploma framework, which might have marked an end to GCSE and A level qualifications, was not fully adopted amidst confusion about what government wanted from 14-19 education (Tomlinson, 2005, p. 152). A key feature of the Tomlinson proposals was the introduction of functional skills (including mathematics) in all proposed 'lines of learning'.

The influential Smith (2004) report made many recommendations concerning the 14-19 curriculum including (in brief):

- 4.1 A change in the structure of GCSE mathematics examinations (from Foundation, Intermediate and Higher tiers to Foundation and Higher tiers only)
- 4.2 Double award GCSE
- 4.3 Reduced time on coursework
- 4.5 Incentives to increase participation in post-16 mathematics
- 4.7 Review and rationalisation of a wide range of mathematics and numeracy qualifications and greater integration of ICT in teaching and learning
- 4.9 Review of pattern and style of GCE assessment
- 4.10 Support for most able at GCE
- 4.11 Develop pathway models, pilot and select 'preferred pathways'

Many of Smith's recommendations have been acted upon quickly. The rise of functional skills, with precedence in earlier attempts to develop key/basic skills, comes as little surprise. However, despite attempts to define functional mathematics (e.g. Roper, Threlfall, & Monaghan, 2006), and a growing literature on quantitative literacy there is still diversity of opinion amongst a range of stakeholders on what functional mathematics is. This seems to be important as qualifications are now being developed to assess functional mathematics skills. Indeed, before the piloting and evaluation of these qualifications has been completed they are being taught as part of the first phase of diplomas which began in September 2008. But the skills deficit is not the only problem that is being addressed in the current pilot. The Smith report explains that "perhaps the greatest concern to many respondents to the inquiry...has been the dramatic decline in A level mathematics entries since curriculum 2000 was introduced".

Recommendation 4.11 paved the way for developing new 14-19 mathematics pathways that might help to redress the long-term concerns about declining mathematics participation. Here was presented the opportunity to rethink mathematics education, creating something appropriate for the needs of learners in the 21st century. Various stakeholders take quite different views and this can be seen in the development of the 14-19 mathematics pathways.

Phase I

In 2005 QCA contracted two teams to develop pathways models: the University of Leeds and Kings College London/Edexcel developed independent models for mathematics pathways.

The KCL/Edexcel model deliberately retained a simple pattern of two GCSEs, A levels and functional mathematics for all pre-18 but had a distinctive emphasis on mathematical modelling and the use of ICT, both within the course and in the assessment of mathematics. Functional mathematics would stand alone and be assessed by a computer-mediated test with portfolio assessment being proposed for trial at all levels. They suggested two models for the GCSE awards; "Model A proposes that both papers cover a mixture of Applications, Skills and Techniques. Model B proposes that GCSE 1 looks at the applications of mathematics across the grade range and GCSE 2 focuses on pure mathematics and an appreciation of the nature of mathematics and mathematical thinking" (July 2006, p.4). They recommended that both GCSEs be required to start the A level. A level is recommended to reduce from 6 to 4 units (similarly AS from 3 to 2): "all units will incorporate a modelling philosophy which, as at GCSE level, would involve a high degree of computer use" (p.5). This model was designed to engage students and build learner confidence through an emphasis on modelling and a belief that some big mathematical ideas are simply worthy of inclusion in the curriculum.

The Leeds model aimed to develop "a curriculum and assessment structure which would encourage more students to study mathematics beyond compulsory schooling, engage

and motivate students and provide students with a mathematically challenging experience within their capabilities" (p. 2.7). Whilst looking similar to the KCL/Edexcel model at GCSE level (with some notable exceptions, e.g. the removal of some statistical content), Leeds proposed a far wider range of components after KS4. The Leeds model included Free Standing Mathematics Qualifications, an A level in Use of Mathematics as well as a GCSE Use of Mathematics. These would be renamed supplementary mathematics modules and some would be designed to meet the mathematical requirements of other A level courses and diploma routes. For Leeds, functional mathematics should also be assessed separately and would consist of a competence and functionality element. They recommended a reduction in A level units to 4 and that greater emphasis should be placed upon proof.

Phase II

Following the development of the two models QCA contracted two awarding bodies (OCR and AQA) to develop various aspects of the Phase I models. AQA are developing assessments based upon the Leeds model and continue to work with the Leeds team. OCR are associated with the KCL/Edexcel model but have not been able to fully implement the recommendations of the Phase I team. This is due in part to the quite significant shift in pedagogy that would be required to implement this model. In addition, the challenges of using ICT have proved problematic. Both contractors designed and ran trial qualifications in 2006-7. These included functional mathematics being a component of GCSE Mathematics which was not the recommendation of either of the Phase I teams. The government policy is that functional mathematics (at Level 2) will act as a hurdle to achieving a higher grade in GCSE mathematics for awards from 2012. The aims for the GCSE in *additional* mathematics were not finalised in the two Phase I models and this lack of clarity has continued into the development of the pilot GCSEs. Currently the two GCSEs are assessing the same programme of study. The second GCSE has a greater emphasis on mathematical thinking and problem solving which reflects changes in the new programme of study for KS4 in which mathematical process skills (including mathematical thinking and problem solving) have a prominent position.

From trials to pilots

In examination terms the difference between a trial and a pilot is that in a pilot students receive examination grades and certificates, as they would for a normal national examination, whereas in a trial students are entered for both a trial examination alongside the current examination and may be awarded their best result. In the summer of 2007 the trials ended and a formal 'pilot' investigation began. At this point QCA announced that functional mathematics would no longer be accredited from the GCSE qualification but would be a stand-alone award. Given the shifting political context in which a project like this is taking place it is to be expected that changes occur that impact on the trial/pilot process. This particular change presented the awarding bodies with two challenges. The first issue was the willingness of centres to participate in the piloting process, although this policy shift did not have a major impact upon participation in the pilot. The more critical consequence is the delay in piloting models that can directly influence the development of qualifications to be implemented in 2010. The decision to change GCSE Mathematics can only be worked through in 2008-9, shortly after which awarding bodies will need to have specifications ready for distribution to examination centres. What seemed like a long period of time for developing pilot qualifications suddenly looks very compressed.

GCSE consultation

QCA have recently undertaken a national consultation on the GCSE subject criteria (the rules that awarding bodies must adhere to when designing specifications). Despite the original intention to have two GCSEs the consultation was concerned with the subject criteria for GCSE mathematics only, together with some questions about the nature and purpose of a second GCSE. This situation arose for a number of reasons but one explanation for the delay in consulting on a second GCSE is due to the fact that different

influential stakeholder groups have different visions for a second GCSE. This was illustrated clearly in the last 12 months in the wide ranging estimates for participation in *additional* mathematics (from the 'most or all' of the Sainsbury Review and down to 50% as reported by government sources at a conference in February 2008)

3 Evaluating the Mathematics Pathways Pilots

Having outlined the background and recent progress of the pathways project we now summarise our approach to evaluating it. Our study is an independent research evaluation involving a large team of researchers from three English universities as well as a team of consultants. The full evaluation study, which is being conducted over three years involves four strands:

- A. Management
- B. School based case studies and surveys
- C. Liaison with professional bodies and stakeholders
- D. Scrutiny of assessment items, qualifications and pathways

Our remit from QCA was to respond to a wide-ranging list of questions which are synthesised under the following four headings:

Meeting students' needs

Central to questions in this theme is the overall question of "how do the proposed new qualifications/pathways appear to meet the needs of students?" Here we see the qualifications as being distinct from the experiences that result: the focus of this theme is therefore the "new" mathematical experiences for learners that result from the pathways and qualifications being put in place. Whilst much of the work of the pathways teams focuses on curriculum and assessment structure, questions in this theme attempt to investigate their fitness for purpose in terms of what learners and their teachers believe are the (developing) needs of students.

Providing pathways and progression

We are investigating whether distinctive and coherent pathways that provide suitable progression routes in mathematics will ultimately result from the project. In this theme therefore, again we focus not on assessment and qualifications, but rather the pathways, and therefore likely trajectories, in and through mathematics available to students.

Assessment

Here we focus on the more tangible work of the pathways projects: that is, the assessment and qualifications that they provide. Our focus in case study schools/colleges is therefore on how this impacts directly on learners and teachers. Additionally this theme allows us to investigate issues such as teachers' perceptions of standards, influences on teaching and so on.

Manageability of implementation

The introduction of the new pathways in mathematics will result in major challenges for those implementing them, not only mathematics teachers but also for a school's or college's curriculum managers. For example, it is likely that they will be offering a wider range of qualifications at all levels in mathematics than previously: how will this be managed?

Strand B: fieldwork and surveys

With around three hundred centres being involved in piloting the pathways qualifications there are hundreds of teachers and many thousands of students from whom we are collecting data. A large proportion of these candidates are doing GCSE (and functional mathematics) at KS4. We are including a range of students and centres in our fieldwork,

both at GCSE level and in exploring GCE, FSMQs and Use of Mathematics. The fieldwork amongst these groups will consist of:

Case Studies

We expect to visit between around 200 centres during the course of the EMP evaluation. Some of these will be longitudinal studies in which we follow a cohort of students through a piloted programme. Others will be short case studies in which we will obtain a broader picture with regard to issues raised in the long visits and through other channels. These case studies include classroom observation, focus groups and semi-structured interviews with a range of teachers and managers. Centres will be sampled from all awarding bodies, school types and regions of the country, covering a range of attainment and social characteristics. This aspect of the work accounts for a significant part of the time and cost of the evaluation and in our view is crucial in terms of revealing detailed evidence from pilot centres about the impact of the pilot on classroom practices, student learning and progression, and school and college policies and strategies as they respond to the opportunities, with which the pilot provides them.

Survey work

We have established and will maintain involvement across centres in various ways and collect further data through on-line surveys, the first of which was carried out in the Spring of 2008.

Strand C: liaison with professional bodies and stakeholders

There is a very wide range of stakeholders that will be affected by the proposed changes to 14-19 mathematics and we are seeking their views throughout the evaluation process. These stakeholders include representatives from professional associations, research organisations, national advisory bodies, various centres concerned with mathematics education, Higher Education, industry and business representatives, and so on. Through existing relationships and newly developed links with these groups we will be seeking feedback on the pathway process throughout the project.

Strand D: scrutiny of assessments, qualifications and pathways

This strand is vital to the evaluation. Here we are considering how assessment items and qualifications are constructed and combine to form pathways. We have developed instruments to analyse qualifications and the extent to which they meet their desired aims. This will enable us to explore the potential of the piloted pathways to tackle the problems of mathematics education that they were designed to address.

4 Some Reflections on our Evaluation of Examination Pilots

The need to Address Complex Educational Issues in the Pilot

Although on the face of it our study is evaluating 'examination pilots', the issues at stake relate to curriculum, pedagogy, student learning and motivation. The fact that awarding bodies are able to successfully run pilot examinations in mathematics does not in any way address the question of whether such examinations are better or worse (from the point of view of their influence on student learning) than the alternatives, which they might in time replace.

Our evaluation has necessarily had to involve a rich variety of forms of data collection including in-depth case studies of schools involved in the pilot, teacher interviews and questionnaires as well as consultations with numerous major stakeholders. In this respect our evaluation is concerned with complex questions about the quality of students' learning of mathematics and the pathways that they can follow through successive years in education to develop their potential in the field of mathematics. Such questions

require very detailed fieldwork and require examination pilots to address issues of classroom practice and the impact of change over extended time periods. The statistics arising from the pilot examinations themselves cannot address questions such as these, even though they may in some cases shed some light on how effectively the examinations themselves are functioning.

Coping with Policy Changes during the live of pilot

The pace of change in mathematics teaching and learning in England is such that the pilot will not run its full length before further changes are made. As a result of the time taken to write, review and sign off examination papers, both in 2008 and 2009 the pilot GCSE mathematics examinations being taken by students in the pilot schools will be 'out of date' in relation to changes agreed and finalised. This situation has a considerable impact on what can be learnt from the pilot. Whilst there is much that can be of benefit in the ongoing evaluation work the project is clearly no longer focused on the original objects of investigation. This can mean that there is a time delay in the evaluation recommendations which may limit the impact of the findings.

Given our experience so far it is entirely possible that there might be other significant changes of direction in the process of developing new mathematics learning pathways. Following the recent national consultation we expect the government to be making decisions about future developments and as evaluators we simply need to be ready to respond to any new scenarios which might arise.

Issues such as these present substantial challenges for the conduct of full scale examination pilots such as the one we are currently evaluating. Examinations such as GCSE generally relate to courses lasting at least two years, so 'examination pilots' of this type for GCSE examinations, which involve syllabus and curriculum adjustments need to be run for at least two years, and are most able to inform decisions about wider scale implementation if they are able to run their full course. However, in our experience such a time period would be a luxury. So anyone planning to conduct future pilots of this type needs to be prepared for the fact that pilot evaluation data will take some considerable time to reach the point where full summative judgements can be made.

Keeping Schools Engaged with the Pilot

However much one values the learning that can occur from 'examinations pilots', such as the one we are evaluating, the successful implementation of such an experiment depends crucially on the willingness of awarding bodies, schools and students to participate in it. We have seen how this willingness can be tested by change. For example, if schools perceive that the pilot could potentially disadvantage their students then they might no longer be willing to continue to participate.

Another issue arises here for those intending to conduct thorough 'examination pilots'. It is clearly highly desirable for the participants in such a pilot to be working under conditions that closely resemble those that would exist if the examinations being piloted were to be made available to all students. This generally means that examinations pilots need to be run as 'live' examinations, which lead to equivalent examination results to normal examinations. In such circumstances students and school agreeing to participate in such a pilot need reassurances that they will not be any way be disadvantaged by the participation in the pilot.

5 Concluding Discussion

The Cabinet Office Strategy Unit report on pilots (Government Chief Social Researcher's Office, 2003) makes a number of interesting recommendations about the future use of pilots in government decision making. These include:

- A pilot should be undertaken in a spirit of experimentation. So, if it is clear at the outset that a new policy and its delivery mechanism are effectively already cast in stone, a pilot is redundant and ought not need to be undertaken
- The full-scale introduction of new policies and delivery mechanisms should, wherever possible, be preceded by closely monitored pilots
- Once embarked upon a pilot must be allowed to run its course
- Independence is critical, pilots must be free from real or perceived pressure to deliver 'good news' and be designed to bring out rather than conceal a policy's imperfections
- Methods matter. A poorly conceived or poorly specified pilot may be worse than no pilot at all."

The current UK mathematics examination pilot, which we have discussed as a particular example within this paper, measures up fairly well to that guidance. The only real difficulty has been in relation to policy changes occurring during the life of that pilot. It certainly stands as a rare example of a long term well-resourced examination pilot with a serious independent evaluation, which is addressing curriculum and pedagogy issues as well as assessment matters.

In conclusion therefore we argue that given the difficulties involved in creating good educational examinations, which fulfil both assessment and curriculum objectives, the concept of piloting new examinations on a realistic sample population of students, schools and colleges is a highly desirable one. Much can be learnt from well designed and properly conducted pilot studies, as the Government Chief Social Researchers Office (2003) concluded. However, there is a need to be both conceptually clear about what is expected from an 'examinations pilot' and to face the implications of organising and properly evaluating pilots. These probably need to run for several years so that trial examination syllabuses and assessment arrangements can be piloted in circumstances which resemble those that will be encountered if the ideas being piloted are to be implemented on a larger scale.

The current evaluation of new mathematics qualifications within the UK is a very ambitious 'examination pilot' involving two major awarding bodies and a large number of pilot schools and colleges. Our evaluation of it in turn had been designed in such a way as to allow us to address some very fundamental questions about student learning and progression, and by the end of the three years of the project we will be able to assemble a substantial body of evidence upon which to base our conclusions. All of this will result from a period of planning, reflection and experimentation going back over at least six years. This is highly unusual in a world of educational decision-making, which is often less evidence-based and sometimes driven more by gut instinct, political imperatives, and crisis management.

In this paper we have attempted to open a discussion about 'examination pilots'. We believe that they can have an important role in guiding education towards improved systems for teaching, learning and assessment and we hope that others will be able to share other examples and perspectives on this critical topic. In essence a true 'examination pilot' is ideally conducted in a setting where key decisions will be informed by the results emerging from the pilot. This condition demands a timetable of decision making lasting years rather than months, and if such a pilot involves the teaching of revised curricula as well as revisions to assessment methods then adequate time needs to be allowed for teachers and students to adapt to the new arrangements. Finally 'examination pilots' are unlikely to reveal in depth evidence unless steps are taken to

evaluate their wider impact on teaching and learning. Assuming that examination results and statistical analyses of examination marks will reveal the effectiveness of the pilot examinations is likely to be a dangerous strategy in most cases where assessment effects are most effectively addressed through more detailed insights into teaching and learning improvements.

References

Government Chief Social Researcher's Office (2003) *Trying It Out. The Role of 'Pilots' in Policy-Making*. London: Cabinet Office Strategy Unit.

Roper, T., Threlfall, J., & Monaghan, J. (2006). Functional mathematics: what is it? In A. Noyes & H. Povey (Eds.), *Research in Mathematics Education Volume 8* (pp. 89-98). London: BSRLM.

Smith, A. (2004). *Making Mathematics Count*. London: The Stationary Office.

Tomlinson, M (2004) *14-19 Curriculum and Qualifications Reform: Final Report of the Working Group on 14-19 Reform*. London: Department for Education and skills.

Tomlinson, S. (2005). *Education in a post-welfare society*. Maidenhead: Open University Press.