

RESEARCH ARTICLE

Seeking authenticity in high stakes mathematics assessment

Pat Drake^a, Geoff Wake^b and Andy Noyes^c

^a *School of Education and Social Work, University of Sussex, UK;* ^b *School of Education, University of Manchester, UK;* ^c *School of Education, University of Nottingham, UK.*

Keywords: Realistic mathematics assessment; mathematics as a human activity; authenticity.

Abstract

This article derives from a scrutiny of over 100 national secondary mathematics examination papers in England, conducted as part of the Evaluating Mathematics Pathways project 2007-2010 by a team of eight researchers. The focus in this article is of the extent to which mathematics assessment items reflect and represent the current curriculum drive for increased mathematical applications in the curriculum. We show that whilst mathematics is represented as a human activity in the examinations, peopling assessment items may serve actually only to disguise the routinised calculations and procedural reasoning that largely remains the focus of the assessments, with the effect that classroom mathematics remains unchanged. We suggest that there are more opportunities for assessment items to illustrate mathematics in use, and we draw attention to ways of assessing mathematics that allow these opportunities to be taken.

Seeking authenticity in high stakes mathematics assessment

Introduction

This article discusses the struggle to embed more realistic contexts into assessment (e.g. Burkhart, 2007), and the difference between the rhetoric and the reality. Rather than consider the effect of such assessments on student performance, in this contribution we ask what challenges these items bring to the process of writing them. We show that in national examinations both individual items and entire assessments are frequently designed with some implicit or explicit ‘relevant’ application. These items are described by various words, and used seemingly interchangeably across stakeholder groups: functional, real life, realistic, authentic, situated, in context, pseudo-contextualised or artificial; and there may be others. Yet, what is also apparent is that these terms mean different things to different people. So through showing some ways that ‘authentic’ mathematics assessment seems to be represented in practice this article also raises a question about language and how meanings of term/ideas vary.

The research reported here has been conducted as part of a national three-year evaluation, commissioned by the Qualifications and Curriculum Authority^j in England of pilot qualifications in mathematics being developed by two awarding bodies. The pilot qualifications are intended to develop new mathematics learning pathways for 14-19 year olds in England, in order to extend participation in mathematics at all levels as strongly advocated in the drive to develop capabilities in science and technology (Smith, 2004; Roberts 2002; Gago, 2004).

The evaluation is addressing two questions:

- What is the likely impact of the proposed qualifications on take-up of mathematics at all levels, particularly post-16, including candidate engagement and confidence?
- Do the benefits of a new system lead to sufficient gains which justify replacing current provision?

The long term decline in pre-university mathematics participation in England has been well noted (Roberts, 2002; Royal Society, 2008; Smith, 2004) and a major review by the QCDA (Matthews & Pepper, 2007) has highlighted a common view in the UK, namely that pre-university (Advanced or A level) mathematics is largely for the ‘clever core’. In addition we know that this cohort is differentiated by gender, ethnicity and class (Mendick, 2005; Noyes, 2009) and that schools have varied impact upon participation (Noyes, 2009) and attrition (Noyes & Sealey, 2009). Around one tenth of each annual school cohort of 16 year olds currently continues to study A level Mathematics (~60,000) but the Government has recently announced their intention of raising this to 80,000 which is a very challenging target.

In addition, it is reported (Moser, 1999) that around 40 percent of adults have some problems with numeracy. Adults without mathematical and numerical competency are known to be disadvantaged, as people with numeracy skills are more likely to be in employment, and earning more at work (see for example Vignoles et al, 2008). A cycle of increasing disadvantage ensues, for adults in employment have more opportunity to practise numerical and reasoning skills when employed at work, and have more access to relevant training. This makes it inherently more difficult for

young people leaving school without basic mathematical competency to ever make up for the disadvantage.

An agenda focusing on functional skills places developing confidence and competence in using mathematics in a wide range of contexts at the heart of the drive to enhance mathematical capability and participation. In developing pathways that might better serve a wider population of students, a number of attempts have been made to develop mathematics assessment, at all levels, that might be more motivating through a clearer focus on applications. These newly emerging assessments might be expected also to serve as a lever for curriculum change that will help develop functionality as well as having the benefit of preparing students more effectively for the type of mathematical activities expected in international comparative studies such as PISA (2006).

Research into developing 'Realistic Mathematics Education' (e.g. Presmeg & van den Heuvel-Panhuizen, 2003) continues to promote the development of 'a school mathematics curriculum that is grounded in the experiential reality of the learners.' p. 1.

Calls for a more realistic mathematics curriculum are contemporaneous with convincing evidence that in England and Wales schools are driven by the need for students to perform well in mathematics and English at the end of compulsory schooling, in fact to achieve grades A* - C in both subjects at GCSE. The impact of this is discussed fully elsewhere (Noyes et al 2008, 2009). Here we draw attention to the obvious inference that in such a political context, for a realistic mathematics curriculum to have any purchase in schools, it must be tested so that these schools can demonstrate success with it. These drivers suggest that we should expect to see a

greater degree of real-life context in mathematics test items to introduce functionality not only into the examination papers, but into the classroom as well.

Studies that consider the the demands of contextualised assessment items on students have been reported by, e.g. Cooper and Dunne, 1998; Cooper and Harries, 2003; Cooper, 2007. Cooper and Harries (2003) usefully present a typology that differentiates between three types of ‘realistic’ mathematics test items. First, those intended to require no extra-mathematical considerations. Second those intended to require particular extra-mathematical considerations, for example realistic situations such as the number of buses needed to carry a particular quite large number of people (where people outnumber seats on a single bus). These require candidates to round up the remainder so as to have a whole number of buses i.e. to provide a ‘realistic’ answer, but not to consider further options, such as people finding alternative modes of transport. Thirdly, problems that are intended to require general extra-mathematical considerations, for example providing alternative answers based on considerations that people in the situation might think about, and asking candidates to explain these. It is argued that both primary and secondary school students are capable of addressing this third type of problem, individually and under test conditions, and demonstrate powers of extended reasoning when doing so. However, as shown in earlier work by Cooper and Dunne (1998), when the item is presented as the second type, candidates may be misled by so-called realistic contexts into applying their everyday knowledge, instead of limiting themselves to the particular mathematical knowledge required for solving the problem. This mistake is less likely to be manifested by high attaining students, who recognise that the context is irrelevant to the mathematical problem on which they are being tested. This means

that the very students whose engagement is being sought through realistic test items are the very candidates who are disadvantaged by them in the examination.

The drive for 'realistic' items suggests that we might find, if we looked, examples of test items constructed possibly in one of three ways identified above, and there is also the potential for new types of assessment not seen previously.

Methods

Assessments are being piloted at three levels, and it would be fair to say that at each level there are also two tracks. A 'traditional' track is represented at the end of compulsory schooling by GCSE, and post compulsory by Advanced Mathematics and Further Mathematics. An 'applications' track that emphasises functional mathematics is represented by qualifications introduced in 2000: Functional Skills and Free Standing Mathematics Qualifications (FSMQs). Level 3, the highest level, represents mathematics usually studied by students choosing to pursue the subject in the post-compulsory phase, i.e. post-16. Level 2 represents student achievement at grades A*-C in GCSE. GCSE is not an age related qualification, although the vast majority of candidates are aged 16. Level 1 is equivalent to GCSE grades D-G. In England school performance is measured by the proportion of students achieving five GCSEs at level 2 including English and mathematics. However, there are also Level 1 FSMQs, and these are proving popular as providing relevant examinations for students at this early level, with level 2 FSMQs providing, during the pilot, an alternative for students who fail to achieve grade C at GCSE and would otherwise have to resit. The numbers of candidates resitting GCSE mathematics is very high (approximately 100 000 p.a) because a grade C is the entry requirement for many academic, vocational and professional courses (eg teaching and nursing). Thus what is

on offer is very complex, and in the pilot qualifications alone, since June 2008 there have been over 70 possible pilot mathematics examinations available to centres involved in the development. This is in addition to the existing suite of over 30 current qualifications in mathematics, and five others in statistics, but statistics is outside the pilot, and in any case accounts for relatively a very small number of candidatesⁱⁱ.

Table 1: Pilot qualifications

Level 1	Level 2	Level 3
<p>Mathematics GCSE grades D – G (Foundation and Higher (grade D only))</p> <p><i>Additional Mathematics</i> GCSE grades D – G (one untiered and one tiered:Foundation and Higher (grade D only))</p>	<p>Mathematics GCSE grades A* - C (Foundation (grade C only) and Higher)</p> <p><i>Additional Mathematics</i> GCSE grades A* - C (one untiered and one tiered:Foundation (grade C only) and Higher)</p>	<p>AS and A level Mathematics</p> <p>AS and A level Further mathematics</p>
<p>Functional Skills</p> <p>Free Standing Mathematics Qualifications</p>	<p>Functional Skills</p> <p>Free Standing Mathematics Qualifications</p> <p>GCSE Use of Mathematics (comprises two FSMQs and functional skills)</p>	<p>Free Standing Mathematics Qualifications</p> <p>AS level Use of Mathematics AS (comprises two FSMQs and an Algebra unit)</p> <p>A level Use of Mathematics (builds on AS and includes calculus, portfolio and mathematical comprehension units)</p>

One layer of the evaluation is item-level scrutiny of assessment items in both pilot and current examinations in mathematics, and this is being undertaken by a team of eight researchers, all experienced mathematics educators. Each examination is mapped, at item level, to a framework that identifies across key dimensions: structure; content; process skills; task type; resources (for more detail please see Wake et al, 2009). To increase reliability, this analysis is conducted independently by two researchers.

Since June 2008, the researchers have scrutinised over 100 question papers at levels 1, 2 and 3. The authenticity of mathematics test items emerges from ‘task type’ dimension of the scrutiny. The following criteria were used:

- Pure – the question has no context other than that of mathematics itself.
- Artificial –whilst a context is introduced it is not authentic in that a candidate would not use mathematics to solve the problem in the way suggested.
- Authentic – the context is something that a candidate could possibly engage with in their day-to-day life and use mathematics in the way the question demands.

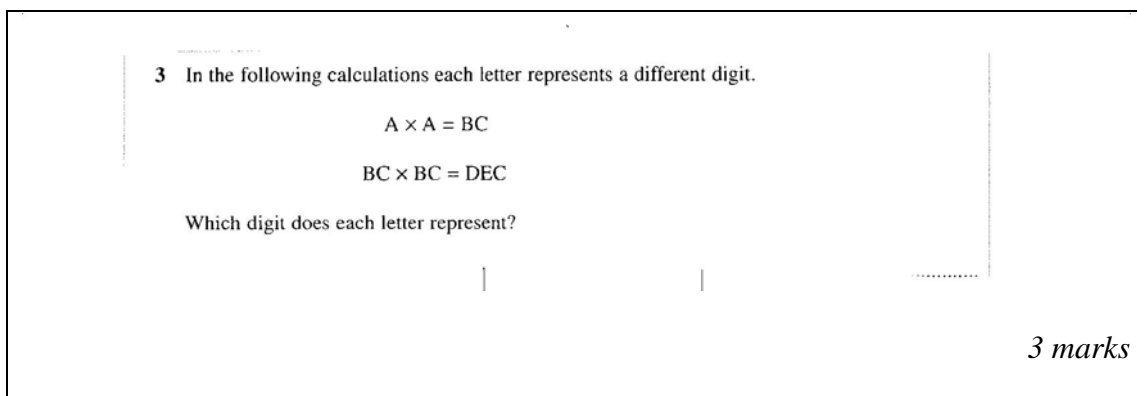
These categories might be seen as loosely corresponding to the Cooper and Harries’ classification introduced earlier, although their typology was generated through studies of pupil performance; and our scrutiny has not considered this, but focused solely on the assessment items themselves. We decided to work with the ideas that ‘pure’ items require nothing in the way of consideration outside the mathematical scenario presented in the question; that ‘artificial’ items require candidates to forego existing real life knowledge, but to use and apply mathematics generated from a

context presented in the examination. We found it more difficult to agree about ‘authentic’ items, and this article reflects that struggle. So for instance it became clear as the scrutiny progressed that ‘pure’ questions are, in mathematical terms also authentic, in so far as these act as preparation for further study of mathematics. We found very little that we could all agree was wholly authentic, but recognised that certain types of assessment facilitates genuine considerations of real life activity and application more than others do.

Examples

The scrutiny work shows that overall, mathematics assessment includes a substantial number of items that signal mathematics as a human activity, used by people in the real world. The examples presented here are chosen explicitly to illustrate the way that the researchers determined how to categorise assessment items; they are also relatively brief. They do not necessarily reflect the tenor of the examination paper from which they are drawn, but have been selected to indicate some of the challenges facing item writers, rather than being typical of the examination as a whole. The examples are all at level 2 or level 3, and drawn from examinations in summer 2008 and summer 2009.

Pure



3 In the following calculations each letter represents a different digit.

$$A \times A = BC$$
$$BC \times BC = DEC$$

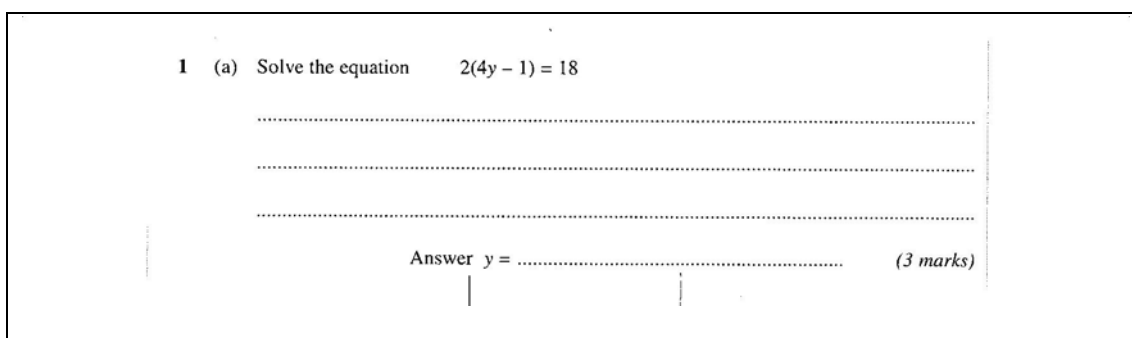
Which digit does each letter represent?

3 marks

Figure 1

There is no context for this GCSE item at level 2 (Figure 1), which requires candidates to reason analytically. There is no expectation that the candidate explains their reasoning, nor extends or extrapolates to a more complex example. It would be difficult to provide an explicit routine for students in advance of the examination for solving this problem, but the appearance of it in the examination signals to teachers that understanding of number theory and skills of deconstructing unexpected representations may be required. This item models a style of mathematical reasoning that students at higher levels are expected to adopt in their studies of mathematics, and so can be considered to be authentic as well as being classified as ‘pure’.

The following GCSE level 2 question (Figure 2) asks for a routine algebraic manipulation.



1 (a) Solve the equation $2(4y - 1) = 18$

.....

.....

.....

Answer $y =$ (3 marks)

Figure 2.

This example highlights two of the tensions inherent in writing ‘pure’ mathematics questions. Firstly algebraic understanding is most frequently tested through questions that require only routine procedural skills. The current subject criteria for mathematics (Ofqual, 2007) expect, for manipulative algebra, weightings of a minimum of 6% of the foundation tier and 22% on the higher tier, and although the assessment of number and algebra together accounts for just over half of the examination (50-55%), this is also weighted so that in the order of 20% of the assessment is algebra on the foundation tier, and 30% on the higher tier. Thus one

would expect to see both quantitatively and proportionately more routine algebraic manipulation on the higher tier, arguably at the expense of other algebraic skills, and the scrutiny confirms that this is the case. Whilst there are relatively few questions of the above type: pure, algebra, solved through routine manipulation at levels 1 and 2, they dominate in level 3 traditional mathematics courses (Noyes et al, 2008, 2009 ; Wake et al, 2009).

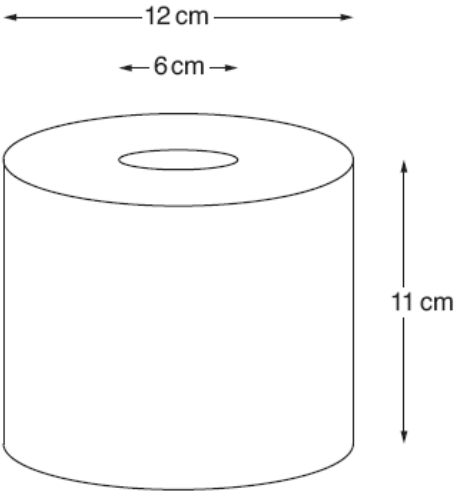
Second, whilst this question looks ‘pure’, there is a context, though as with the previous item this is outside the immediate examination. Students who continue with mathematics into the post-compulsory phase to pursue traditional mathematics courses need skills in algebraic manipulation. This is a regular refrain of stakeholder groups, also confirmed in our reports. This fact introduces a tension into the curriculum pre and post-16, and teachers in each phase are not in agreement. Post-16 teachers in pilot centres characteristically bewail the need to teach all these ‘boring procedures’ at the expense of interesting mathematics because algebraic procedures are not taught earlier, whereas teachers in the compulsory phase tend to resist teaching too much routine algebra as number and algebra comprise just 20% of the national curriculum for mathematics, and as explained above, it is not expected to dominate the assessment. However, according to post-compulsory teachers of traditional mathematics this is insufficient preparation.

Artificial

The question that follows (Figure 3), albeit amusing, illustrates a distinctive feature of mathematics assessment items. An attempt is made to dress up, by introducing a pseudo real life context, required mathematical procedures, in this case, demonstrating skills of drawing a plan and an elevation, and at the higher tier,

estimating a fairly standard volume. However the context is entirely unrealistic for school students (and it would be interesting to find a context where this is realistic). A problem solving element, even more unrealistic, is also introduced for higher tier candidates, that of estimating the number of sheets on the roll.

7 The diagram represents a toilet roll.



(a) Draw a full size accurate side elevation of the toilet roll. *2 marks*

(b) Draw a full size accurate plan view of the toilet roll. *2 marks*

Figure 3.

This is a level 2 question from the GCSE foundation tier non-calculator paper. The higher paper in the same year goes on (Figure 4):



(c) (i) By taking π equal to 3, find an estimate for the volume of paper on the roll.

(ii) Each sheet on the roll has a volume of 4 cm^3 .

Work out an estimate of the number of sheets on the roll.

2 marks

Figure 4.

Fortunately in this example, most of the question posed is so far outside everyday experience that candidates are unlikely to fall into the trap outlined by Cooper and Dunne (although some may have specialised external non-mathematical knowledge that they bring to the last part!). Incidentally, one wonders whether the ‘joke’ in this item is intended for the candidates, or for teachers searching for interesting and illustrative examples with which to liven up future revision lessons.

On the other hand, the following example (Figure 5) is also artificial, though signalling to candidates that mathematics is useful and applied in real life, so our team would consider this item to be part artificial and part authentic. Questions about carpets in mathematics assessment are fairly common, as the rectilinearity lends itself to testing area and perimeter, either numerically or through algebraic modelling. This is one that presents the carpet problem as something that the candidate might address within the context of their home.

11 Michelle is buying carpet and a wallpaper border for her bedroom.
Her bedroom has floor area 20 m^2 and perimeter 18 m.
She has chosen a carpet costing £15 per square metre.
Rolls of border are 5 m long.
She has £350 to spend.

After paying for the carpet, which of these borders is the most expensive she can afford?




	Smiley face £7.00 per roll
	Hearts £10.00 per roll
	Trees £13.00 per roll
<i>6 marks</i>	

Figure 5.

On first sight this question appears to be realistic, but on closer inspection we realise that it is of the second type identified by Cooper and Harries, namely that the realism relates to the specific mathematical considerations, with the candidate required to forego all other ideas they might have about carpeting their bedroom, particularly the fact that the area is not the only consideration when ordering carpet – the dimensions of the room matter as well as carpet being made in particular widths. There is also an ambiguity here as well, for it is clear to the mathematical reader that the borders are borders of wallpaper, with this being made explicit by the introduction of ‘a’ into the opening sentence:

*Michelle is buying carpet and **a** wallpaper border for her bedroom.*

However the storyline suggests that borders may also be carpet. And so, unless the candidate ignores the carpet element, they will get the question wrong, although they will have undertaken a more complex calculation that requires more reasoning.

We include one more example (Figure 6) below from level 2 chosen to illustrate mathematics being useful or necessary for future careers, and so also signalling a potentially realistic application of mathematics. However, the data looks as if it may have been invented, and it would be unlikely that an accountant was actually taking this examination, so our scrutiny would place this more in the artificial category. In this example the candidates are not required to forego other non-mathematical considerations, and so it can also be seen as pure routine calculation of a moving average, with some interpretation and communication of the result taking the candidate back into a world of accountancy.

In fact in 2009, across all of the pilot GCSEs there were only a few such explicit examples of professional use of mathematics: this one about an accountant; a question about the dimensions of an aircraft wing and another about the frame of a bicycle, indicating possible links with engineering; a third about planning an octagonal school building, with the story suggesting implicitly that the candidate is an architect; and a couple more introducing characters called Ali and Ben who were respectively estimating the height of a flagpole, and estimating the height of a building. Although the illustration suggested that Ali was not yet an adult, it would be possible to link these questions to occupations such as surveying or construction.

12 The table shows the number of visitors to an art gallery.

Year	2007				2008			
Quarter	1	2	3	4	1	2	3	4
Visitors (thousands)	9	14	16	13	13	16	20	14

The art gallery's accountant calculated the 4-quarter moving averages of the numbers of visitors.

(a) Complete the list of 4-quarter moving averages.

.....
 13000 14000 14500 _____ 15750

[1]

(b) In his report, the accountant said, "The number of visitors is increasing."

Was he correct?
 Give a reason for your answer.

_____ because _____

_____ [1]

Figure 6.

Authentic

1 A driver wanted to rent a car for one week in France. At the time, the cost was 275 euros and the exchange rate was 1 euro = 80 pence.

(a) The driver opted to pay in euros using her credit card which charged a 2.75% fee.

Calculate the cost of renting the car in pounds giving your answer to the nearest penny.
(2 marks)

(b) The cost of renting this car if the driver had opted to use Dynamic Currency Conversion would have been £ 229.35 .

What exchange rate has been used to do this conversion? *(2 marks)*

Figure 7.

This item (Figure 7) is drawn from a level 3 FSMQ Mathematical Principles for Personal Finance, and reflects a topical concern. Candidates facing this question are likely to be at least 17 years of age, and may be older - it would be wrong to assume a school-age audience – and so it is possible that this represents a potential experience

for people taking the examination. It seems that the choice that customers overseas may exercise through dynamic currency conversion could prove to be expensive, making this a real problem.

In this examination, and for all FSMQs at each and every level, pre-release material is made available to candidates two weeks or so before the examination. For this item, the pre-release material is as follows (Figure 8):

<p>Foreign transaction fees</p> <p>When you make a purchase in a foreign country using a credit or debit card you are typically charged a fee of 2.5%, although different banks charge different amounts and a few make no additional charge. When you get your statement at home the total will have been converted into British pounds.</p> <p>An alternative increasingly being used is called Dynamic Currency Conversion. This system recognises that your card is British when you make a transaction overseas and asks if you want to make the purchase in pounds rather than in the foreign currency. This has the advantage that you know exactly how much will be charged in your own currency. However, the overall charge is likely to be greater as the exchange rate used to do the conversion is likely to be disadvantageous, typically incurring an additional cost of 4%.</p>
--

Figure 8.

The pre-release material is a necessary adjunct to the assessment item, as it gives the candidates time to research the context before the examination. Candidates are not allowed to take their pre-release material in with them to the examination, but they are provided with a clean version once they get there.

Pre-release material is used at all levels where functionality or application is the stated aim of the assessment, and may introduce new ideas to candidates, providing an opportunity for genuine problem solving to be undertaken. All pre-release material introduces human (or animal, see below) contexts in which mathematics may represent or be used to model real circumstances, and is provided with the additional

intention that candidates become familiar with the contexts to be used before entering the examination hall. The combination of pre-release material and question paper offers genuine opportunity for exploratory and open, investigative teaching even though the actual assessment item may be asking quite closed questions.

In theory pre-release material provides time for teachers to help candidates navigate what might be asked, what is relevant and what is actually contextual window-dressing. This introduces a problem however; for most schools and colleges are off-timetable in the lead up to examinations and students are on study leave, leaving no teaching time in which to unpack some of the new ideas. In something topical, as in the example above, this will advantage candidates who have heard of the concept through other means. Dynamic currency conversion has only recently been brought to public attention by the popular media – radio and television – with those who are likely to have paid attention being those planning to make overseas purchases using credit or debit cards.

Another potential consequence of lack of teaching time is that the pre-release material becomes trivial or irrelevant, as in the case below (Figure 9) taken from FSMQ Dynamics, also at level 3:

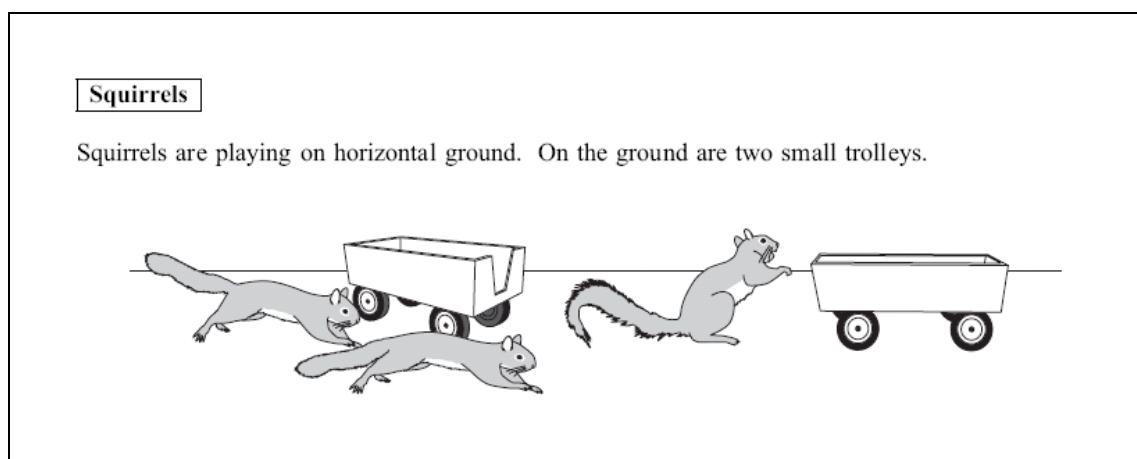


Figure 9.

It is obvious mathematically that the squirrels are unnecessary except as a means of making the trolleys move, and indeed this is the case, as the question shows:

- 2 One squirrel, of mass 0.5 kg, jumps onto a stationary trolley, of mass 1.5 kg. Just before the squirrel lands on the trolley, the squirrel has a horizontal speed of 2 m s^{-1} . After the squirrel lands on the trolley, it and the trolley move together along a straight line in the same direction as the initial motion of the squirrel.
- (a) Find the speed of the squirrel and the trolley immediately after the squirrel lands on the trolley. *(2 marks)*
- (b) The squirrel stays on the trolley. After the squirrel lands on the trolley, the trolley is subject to a horizontal frictional force, F , and stops in 0.1 metres.
- (i) Calculate the deceleration of the trolley while this force acts. *(2 marks)*
- (ii) Find the magnitude of the frictional force. *(2 marks)*
- (iii) Hence find the coefficient of friction between the trolley and the ground. *(3 marks)*

Clearly this question could be addressed by candidates without the benefit of the pre-release material, whose purpose in this case at best signals the need to revise laws of motion – which would seem unnecessary for an examination with the title ‘Dynamics’. These squirrels render the item artificial. Without the squirrels at all it would be pure.

Discussion and conclusion

Mathematics question papers on the whole tend to be presented in generously spaced booklets, with some questions taking up to two pages because of an introductory storyline, diagrams, graphs, and space for answers. Questions frequently introduce a human element, and a scan of pilot GCSE papers in 2009 revealed a wide variety of activities: Mary, Jody, Susan, Fred were all ‘thinking of a number’; Jeff failed his driving test three times; Pete made shelves; Viki set up a mobile phone contract; several named individuals were buying sandwiches, drinks and crisps; yet more

named individuals were saving money and considering interest rates; lots of people were making decisions based on price reductions; Mehdi, Tim and Mr Taylor were driving; and there were various questions involving anonymous cyclists and motorists. There is no doubt that mathematics is signalled as a human activity in all assessments at level 1 and level 2; and also in FSMQs at level 3.

Scrutiny of items has enabled us to begin to differentiate between assessments that promote mathematics in this way, viz:

- Something the candidate ‘ought’ to know about as a citizen, e.g. managing personal finance. The ‘dynamic currency conversion’ has this feel to it, and represents a sub-type of citizen item about learning to protect oneself from losing money, or even from being cheated. There are also items in this citizenship genre about evaluating evidence in a rational manner such as questions asking candidates to determine medicine dose, spread of disease etc. These types of item come with health warnings of cultural and gender bias, but offer potential for truly exploring mathematics as a modelling tool in realistic and relevant situations, predicated on real data and real information. However this type of question is rare.
- ‘Real life’, where assessments are set in the context of activities that the candidate might engage with, e.g. reading timetables, or scheduling tasks. Without pre-release material though it is rare to find genuine data being used, and unfortunately real life is mostly only used for the story line of the problem, as with the example of Michelle redecorating her bedroom. With pre-release material it is possible to offer students at all levels genuine contexts in which to use mathematics.

- Problems arising from genuine parameters (such as cost, time, dimension), but specifically mathematical in the skills, knowledge and understanding expected. Extreme examples of this form are seen in ‘Toilet Roll’ and ‘Squirrels’. These items may amuse and provide material for practising the routine skills that teachers understand underpin most of the examination.
- Problems that specifically address particular mathematics in order to support students in other areas of study, e.g. engineering. Actually this is rarer than we assumed it would be, certainly in the scrutiny of level 1 and level 2 items in 2008 and 2009. There would seem to be some scope for development of this type of item in the compulsory phase.
- Vocational assessments where mathematics may be implicit, or, at higher levels requiring more explicit skills. This is not apparent at all in standard mathematics assessment, but it is intended that these should become evident as functional skills bed into the vocational curriculum currently being developed.
- Problems that are mathematically genuine in so far as requiring candidates to demonstrate skill and aptitude for what is likely to follow in terms of traditional mathematical pathways. These come in different guises: first, problems that test the the ability of the candidate to reason under examination conditions; and second problems that test the candidates grasp and facility of routine procedures.
- Another setting for genuinely mathematical problems uses the pre-release plus examination combination to develop candidates’ understanding of mathematically modelling real situations, usually with a scientific application, such as logarithmic growth. These unashamedly promote mathematics as a high level tool for solving problems. The Use of Mathematics level 3

assessment currently includes a comprehension paper of this type, but there is no evidence of these settings at lower levels of standard assessments, so it is likely that students will meet this approach for the first time when (and if) they progress far enough to work at level 3.

Including people and animals in the assessment items does give mathematics papers a friendly feel, although as some of the examples above show, anthropomorphising mathematics can lead test questions into the realms of fantasy rather than real life; with applications becoming spurious and unreal. We also report elsewhere (Noyes et al, *ibid*) that the majority of mathematics assessment require candidates to demonstrate technical and procedural competence, with, at every level, very little requirement for understanding, analysis, interpretation, representation or mathematical communication. In the few cases where these process skills are required relatively few marks are allocated to them.

If as we assume, what is in the mathematics assessment is closely connected to what the curriculum becomes for teachers and students, more problematic is the further purpose these items serve beyond the immediate testing of candidates. Used by teachers to prepare for future examinations, it is the actual assessment items that help teachers and students recognise the range of mathematical activity that is expected for good performance to be achieved, just as much as the curriculum specification. We have noted (Noyes et al, *ibid*) that teachers in the pilot are consistent in expecting specimen materials to align closely with what actually appears in the examination. The introduction of a human element into this process may make a hitherto unpalatable and boring curriculum slightly more humorous and palatable. It may even

disguise for some candidates for some time the fact of the curriculum focusing on routines and not on problem solving. A new secondary curriculum will become statutory for 14-16 year olds in 2010 and has far greater emphasis on problem-solving, functionality and mathematical thinking. Consequently papers will need to change to incorporate more items that assess this. However, authenticity is not simply a matter of including people into assessment items, as ironically this runs the danger of reinforcing a routinised curriculum, rather than breaking the current stranglehold that procedural mathematics currently holds.

References

- Burkhardt, H. 2007 Mathematical proficiency: What is important? How can it be measured?, in Schoenfeld, A.H. (Ed), *Assessing Mathematical Proficiency*, Cambridge University Press, Cambridge.
- Cooper, B. 2007 Dilemmas in designing problems in 'realistic' school mathematics: a sociological overview and some research findings. *Philosophy of Mathematics Education Journal* 20 (Special Issue on Social Justice: Part 1).
- Cooper, B and M. Dunne 1998 "Anyone for tennis? Social class differences in children's responses to national curriculum mathematics testing", in *The Sociological Review*, 46, 1, 115– 148.
- Cooper, B. and T. Harries 2003 "Children's use of realistic considerations in problem solving: some English evidence", in *Journal of Mathematical Behavior* 22 451–465
- DfEE 1999 *A fresh start: Improving literacy and numeracy. The report of the Working group chaired by Sir Claus Moser*. London: Department for Education and Employment, DfEE.

- Gago, J. M. 2004 *Increasing human resources for science and technology in Europe*.
Brussels: European Commission.
- Matthews, A. and D. Pepper 2007 *Evaluation of Participation in A level Mathematics: final report*. London: Qualifications and Curriculum Authority.
- Mendick, H. 2005 Mathematical stories: why do more boys than girls choose to study mathematics at AS-level in England? *British Journal of Sociology of Education*, 26, 2, 235–251.
- Noyes, A. 2009. Exploring social patterns of participation in university-entrance level mathematics in England *Research In Mathematics Education*. 11, 2 167–183
- Noyes, A., Wake, G., Drake, P. and R. Murphy 2008 *Evaluating Mathematics Pathways* www.nottingham.ac.uk/emp Executive Summary End of Year Report Stage 4 May–December 2008. Full report available from the QCDA on request.
- Noyes, A., Wake, G., Drake, P. and R. Murphy 2009 *Evaluating Mathematics Pathways* www.nottingham.ac.uk/emp Executive Summary Interim Report April 2009 Stage 5: January–April 2009. Full report available from the QCDA on request.
- Noyes, A. and P. Sealey 2009. Investigating participation in Advanced level mathematics: a study of student drop out *Research Papers in Education (in press)*.
- Ofqual, 2007 *GCSE mathematics subject criteria*, QCA/06/2901. Accessed online 28 August 2009 http://www.ofqual.gov.uk/files/qca-06-2901_gcse_maths_2007.pdf
- PISA 2006 *Assessing Scientific, Reading and Mathematical Literacy: A Framework for PISA*

- Presmeg, N. and M. Van Den Heuvel-Panhuizen 2003 'Leen Streefland's Legacy'
Educational Studies in Mathematics, Realistic Mathematics Education
Research: Leen Streefland's Work Continues 54, 1, 1-4
- Roberts, G. 2002 *SET for Success: The supply of people with science, technology,
engineering and mathematical skills*. London: The Stationery Office.
- Royal Society 2008. *Science and mathematics education 14-19: A 'state of the nation'
report on the participation and attainment of 14-19 year olds in science and
mathematics in the UK*. London: The Royal Society.
- Smith, A. 2004 *Making mathematics count – the report of Professor Adrian Smith's
Inquiry into Post-14 mathematics*.
- Vignoles, A., de Coulon, A., and O.Marcenaro-Gutierrez 2008 *The value of basic
skills in the British labour market. Research report*. London Institute of
Education: National Research and Development Centre for adult literacy and
numeracy.
- Wake, G., Drake, P. and A. Noyes 2009 'How exactly does assessment of
Mathematics define student activity?' Paper presented at European Conference
of Educational Research, Vienna, 2009

ⁱ From August 2009 Qualifications and Curriculum Development Agency

ⁱⁱ There are 2 GCSEs, 2 AS levels and 1 A level in Statistics, taken by about 1500 candidates.