Assessing 'functionality' in school mathematics examinations: what does being human have to do with it?

Pat Drake[a], Geoff Wake[b] and Andrew Noyes[b]

[a] *School of Education and Social Work, University of Sussex, UK;* [b] *School of Education, University of Nottingham, UK;*

Abstract

This article analyses aspects of the process of developing 'functional' assessments of mathematics at the end of compulsory schooling in England. A protocol that was developed for scrutinising assessment items is presented. This protocol includes an indicator of the 'authenticity' of each assessment item. The data are drawn from scrutiny of 589 assessment items from thirty-nine formal unseen examinations taken by students aged sixteen, and the article illustrates ways that mathematics is presented in different contexts in examinations. We suggest that currently the 'human face' of the questions may serve to disguise routine calculations, and we argue that in formal examinations connections between mathematics assessments situated in context and functional mathematics have yet to be established.

Introduction

The relationships between schooling and the subsequent contribution that citizens make to the economy is of direct interest to all countries, and forms a major part of a globalised education policy discourse (Rizvi and Lingard 2010). A current preoccupation for OECD countries is the Science, Technology, Engineering and Mathematics 'STEM' agenda which has the objects of both strengthening science and technology in society, and increasing economic growth through strengthening national capabilities in science, technology, engineering and mathematics (Gago

1

2004). The learning of mathematics is central to this agenda, and includes two principal strands: to increase the number of mathematically highly-qualified people in the workforce; and to extend the repertoire of mathematical facility more generally on the part of a wider proportion of the workforce. In England, as is explained in detail later, the collection of mathematical skills that learners use in being effective citizens, is defined broadly as 'functional mathematics' and carries an implication that functional skills include problem solving abilities.

This article results from work scrutinising mathematics school examination items undertaken as part of a project evaluating school mathematics qualifications and addresses the current enterprise of assessing functional mathematics in school examinations (e.g. Burkhart 2007), in particular discussing tensions that arise when mathematics in examinations is put into the context of real life. One of the aims of this scrutiny was to understand the relations between, in mathematics assessments, 'functionality' and the extent to which mathematics assessment is contextualised in, and reflective of, genuine human activity.

An extensive corpus of work by Cooper and colleagues (e.g. Cooper and Dunne 1998; Cooper and Harries 2002) has already considered the effect of 'realistic' assessments on student performance in tests, showing that performance is related to socio-economic status. Cooper subsequently considers the challenges of writing 'realistic' test items, in the light of both the known inequities in students' responses on the one hand combined with pedagogic desires to make mathematics more meaningful (Cooper 2007), arguing that a different type of item, more open to a range of answers is required. Elsewhere Gates examines philosophical questions of social justice inherent in presenting mathematics in human settings, e.g. 'best buy' problems might encourage students to question whether it is 'fair that poorer people cannot afford to buy the more economic large boxes of cereal' (2002, 224).

In this article, a protocol designed for the purpose of scrutinising mathematics test items is presented. Rather than consider student responses, we draw on data obtained by using the protocol to discuss the humanising of mathematics assessment items and show qualitatively that introducing humans into assessments items does not square with the problem solving that functionality requires. It is likely that, as the earlier work has shown, these human contexts for mathematics would serve to distract different groups of students inequitably, encouraging some to 'problem solve' in ways that would not achieve the required answers and marks. However, we also observe that as contextualised mathematics examination items rarely offer problem solving opportunties, equity in performance may not be the most significant concern.

The interaction between what is assessed and what is taught has been discussed extensively (e.g. Hodgen and Wiliam 2006; Stobart 2008), and whilst 'assessment for learning' is generally considered to be an effective way for students to develop understanding and skill, the alternative of learning for assessment, or 'teaching to the test' has the effect of narrowing the curriculum, as Stobart notes (2008, 124), to what is likely to be on the test. Our data show that examination items that represent mathematics 'as a human activity', far from providing opportuntities for demonstrating functionality through problem solving, frequently merely serve to dress up routine calculations and procedures. Thus, contextualising examination items in the way we found actually mitigates against problem solving in mathematics, for it is these humanised, but routine, calculations that effectively become the curriculum for students preparing for the examinations.

The data presented is drawn from scrutiny of 2008 and 2009 examination papers taken in England at the end of compulsory schooling by students generally aged 16. This examination is called the General Certificate of Secondary Education (GCSE), and spans levels 1 and 2 in a national

framework of qualifications that runs from Entry level to level 8[i], where, in terms of higher education, level 6 represents honours degree achievement and level 8 represents doctoral level. GCSE is a national examination in the sense that it is designed to assess, according to government specified criteria, the national curriculum for 14-16 year olds. Candidates sit two or sometimes three examination papers under traditional examination conditions, i.e. questions are previously unseen by the candidates, they work individually in silence, and there is no contributory coursework. GCSE is offered by three main awarding bodies, with schools entering their students for the examinations from an awarding body of their choice. Successful candidates will achieve Grades A* – C and this is considered comparable to other level 2 qualifications. Candidates in the range G – D at GCSE are considered unsuccessful at level 2 but successful at level 1.

Background

Educators and assessors may well agree with a broad and future-oriented definition of functional mathematics such as that provided in the Programme for International Student Assessment (PISA) framework which states that mathematical assessment should:

> focus on real-world problems, moving beyond the kinds of situations and problems typically encountered in school classrooms. In real-world settings, citizens routinely face situations in which the use of quantitative or spatial reasoning or other cognitive mathematical competencies would help clarify, formulate or solve a problem. Citizens in every country are increasingly confronted with a myriad of tasks involving quantitative, spatial, probabilistic and other mathematical concepts.

PISA 2009, 84–85

However, agreeing with the spirit of the definition does not mean that everyone would, in practice, privilege the same modes of assessment in mathematics. What is a useful application for one person may well seem to be an irrelevance for another. So for instance the provision on the one hand of a

sufficiently specific and algebraic curriculum for those likely to participate in advanced study of mathematics; and the wider ranging requirements of an educated citizenry to deal an increasingly data-driven society on the other are in tension. These tensions are not inconsiderable, and they are challenging to reconcile.

A crisis in England

In England, a number of influential reports (including Roberts 2002; Tomlinson 2004; and, most influential for mathematics, Smith 2004) have sought to address a so-called 'crisis' in mathematics education. The concerns are two-fold. First, the Secretary of State for Education has expressed concerns about national placing in PISA (dropping in mathematics from 8th in 2000 to 27th in 2009). Second, a range of interest groups have been concerned about low participation in post-compulsory mathematics (Hodgen, Pepper, Sturman and Ruddock 2010). An apparent change in this trend of post-16 participation has been apparent since 2008, with both numbers of candidates taking mathematics A-level increasing, and the proportion of mathematics entries out of all A-level entries increasing from 7.8% in 2008 to 9.6% in 2011 (Joint Council for Qualifications, 2012). In spite of this improvement, the recent 'Vorderman' report states that still 'nearly half of our students 'fail' GCSE' and that 'only 15% of students take mathematics, in some form, beyond GCSE' (2011, 3). A major review by the Qualifications and Curriculum Agency (Matthews and Pepper 2007) highlighted a common view in the UK, namely that pre-university level 3 mathematics is largely for the 'clever core'. This means that, unlike other subjects, the majority of students taking mathematics at A-level obtained the very highest grades of A* and A at GCSE; in other words, post-16 mathematics is not generally accessible to students other than the highest achievers. In addition we know that this cohort is differentiated by gender, ethnicity and class (Mendick 2005; Noyes 2009); and that there is also a school effect (Brown, Brown and Bibby 2008) with school differences evident in both participation (Noyes 2009) and attrition (Noyes and Sealey 2012).

The Tomlinson report focused on the relationships between qualifications, relevance and participation, suggesting that young people should continue to study relevant mathematics from 14 – 19 years in the context of a qualification structure in which academic study formed one option from a range of vocational routes in which young people were to be encouraged to participate. Recently the Wolf Report (2011), whilst shifting the focus from mathematics per se to vocational opportunity, has emphasised the need for all young people to continue with mathematics until they have achieved success in standard qualifications. Critical of 'functional skills' on the grounds that they require an unreasonable level of contextual knowledge of mathematical applications on the part of mathematics teachers, and wary of embedding mathematical skills in context because they become 'they embed to the point of vanishing' (2011, 170) the report highlights an inherent tension in understanding functionality. Is it about representing real applications of mathematics, or is it about developing process skills of mathematical problem solving? Nevertheless Wolf argues that the GCSE end of compulsory schooling qualifications in mathematics, and also in English become the means of providing young people with the credentials that they need for realistic opportunities at work. Thus, GCSE in mathematics is likely to remain an important gatekeeper to a range of educational, employment opportunities.

The problem of finding a functional terminology

The mix of ideas that lead to mathematics situated in some meaningful or applied context, being described in different ways, such as realistic (e.g. Presmeg and Van Den Heuvel-Panhuizen 2003), mathematically literate (Steen 2001) functional (e.g. Roper, Threlfall and Monaghan 2006), become problematic when terms are used interchangeably, for as indicated above each means different things to different people. Steen (2001), working in the US, shows how developing a consensus is not straightforward, for the working definitions used by international comparative studies exemplify different assumptions. These are brought together under the umbrella 'quantitative literacy' that

addresses the preoccupations of: using quantitative tools; the ability to understand and appreciate the role of mathematical and quantitative methods in world affairs; basic skills; and higher order thinking. Functional mathematics offers an inflection of this debate in the UK.

> The term 'functional' should be considered in the broad sense of providing learners with the skills and abilities they need to take an active and responsible role in their communities, everyday life, the workplace and educational settings. Functional mathematics requires learners to use mathematics in ways that make them effective and involved as citizens, operate confidently and to convey their ideas and opinions clearly in a wide range of contexts.

Qualifications and Curriculum Authority [ii] 2007

This broad working definition implies that functionality means developing problem solving abilities. Functional skills assessment criteria focus on three inherent problem solving and interrelated process skills:

- Representing – selecting the mathematics and information to model a situation
- Analysing – processing and using mathematics
- Interpreting – interpreting and communicating the results of the analysis

Qualifications and Curriculum Development Agency[ii] 2009

As all 16 year olds complete the GCSE mathematics examination intended to assess the secondary school national curriculum, and as the examination criteria also include the functional process criteria above, all successful candidates have ostensibly achieved functionality in mathematics. Functional Mathematics is also available as a separate qualification at levels 1 and 2.

Methods of exploring assessment: developing some analytical tools

The research reported here was conducted as part of a national three-year evaluation 2007 – 2010, commissioned by the Qualifications and Curriculum Authority in England of pilot qualifications in mathematics at levels 1, 2 and 3 being developed by two awarding bodies. The pilot qualifications project was called the 'Mathematics Pathways Project' and ran for five years from 2005 – 2010, with the qualifications intended as a key part of the development of new mathematics learning pathways for 14-19 year olds in England. What was on offer during the Mathematics Pathways Project was complex, and since June 2008 there have been over 70 possible mathematics examinations available to approximately 600 centres involved in the development. This is in addition to the existing suite of over 30 current qualifications in mathematics. The evaluation was tasked with responding to the following two questions:

- What is the likely impact of the proposed qualifications on take up of mathematics at all NQF/QCF 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, on the basis of student experience, manageability for teachers, and the assessments themselves?

This second strand included scrutinising the ways that functional mathematics was included in the GCSE mathematics examinations and focused solely on the assessment items themselves rather than on students' responses to them. The findings reported are generated from this assessment strand of the evaluation, which had a particular focus on the following:

1. The alignment of GCSE with the aims and objectives of including 'functionality';
2. Relations between the curriculum and assessment when working towards GCSE and the likely learning experiences of students.

The analytical framework shown in Figure 1 was developed and used to scrutinise the component tasks of any qualification on an item-by-item basis. (The entries in Figure 1 show how the assessment item shown in Figure 3, 'Toilet Roll' was analysed in the scrutiny.) The PISA framework emphasises the centrality of the problem or task and three key domains which need to be taken into account to understand what is then required of the problem's solver: (i) the mathematical content (and concepts) that need to be brought to bear (ii) the competencies or process skills required to arrive at a solution, and (iii) the situation or context in which the problem is situated. The EMP protocol was a spreadsheet designed so that content matched the English national curriculum domains of Number and Algebra, Geometry and Measures, and Statistics, although in separate categories so as to be able to identify the number, algebra, geometry, measures, statistics and probability content explicitly. The process skills Representing, Analysing, Interpreting and Communicating followed, with a distinction made between analytical reasoning and procedural reasoning. Context of the assessment item was considered as Pure, Artificial and Authentic. Thus the protocol represents a means of describing the functionality of the test items in terms of both the English national curriculum and the more internationally recognisable PISA framework. In addition, the structure of each assessment item was recorded in terms of the number of parts the item included, and the number of marks per part, so that the overall structure of the distribution of marks across items in a single examination could be looked at with a view to evaluating what the examination as a whole required candidates to do. Finally the scrutiny work included making a qualitative assessment as to how difficult an item was likely to be in terms of its complexity and its familiarity to the candidates; and we noted whether calculators, computers and other equipment (such as rulers and protractors) were allowed, needed or likely to be useful.

Insert Figure 1

Over two years 2008 – 2009, ninety-four examination papers were scrutinised as shown in Table 1 by a team of eight mathematics education researchers.

Insert Table 1 here

In this article we focus specifically on the thirty nine GCSE examinations at level 2, represented by 589 examination items in total across two years. Interested in the extent to which the assessment items set mathematics in context as well as in the mathematical content and process skills required we devised context criteria. 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. Authentic items were defined as being when 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 demanded.

This latter category became evidently problematic as items were scrutinised, for making judgments about authenticity with respect to day-to-day life immediately begged the question of perspective: whose life, researcher, teacher, student; and which researcher, which teacher and which student; for naturally all experiences of the everyday are different. Whose everyday experience was it that should be considered? Secondly, as Cooper and collegues (ibid) have investigated from the perspective of student response and explained above, students do not think or behave in the same way mathematically in real life as they are required to think in a mathematics examination to be successful; and thirdly, the examination is itself 'the elephant in the room' providing an intermediary setting for doing mathematics that cannot possibly be the same as the authentic situation in which the mathematics is set. This point is argued more widely also by Gainsburg (2005) who suggests that no

classroom mathematics can ever be 'authentic' as the real life or work context also provides in itself learning about problem solving in situ that can not be replicated elsewhere. We found very little that was wholly authentic in these examinations, but recognised that certain types of assessment facilitate genuine considerations of real life activity and application more than others do, and these are discussed later in the article.

Each examination was scrutinised by two people who arrived at an agreed picture for it. Available to the team were qualification papers, mark schemes and analysis grids provided by the awarding bodies. Scrutineers worked through the test items, recording electronically on the spreadsheet represented in Figure 1 the structure, content, process skills, context complexity and unfamiliarity of each item in the examination. Completing the spreadsheet generated six diagrammatic profiles for each domain in the examination overall. Figure 2 illustrates two, the context and process skills of one examination paper in which nearly 90% of the items were considered to be 'pure', and about 10% of the items 'artificial', i.e. the mathematics was presented in a context, but the context was not necessary for solving the item and indeed might have been an entirely fabricated use of mathematics. In this example the team considered that fewer than 2% of the question parts required the candidates to operate in what could be considered 'authentic' contexts, in order to answer the questions correctly. Figure 2 also illustrates that the same examination required candidates to use routine procedural analysis to obtain more than three-quarters of the marks. More detail of the approach and of the analysis of the examinations can be found in the EMP Stage Interim Report 3 (Noyes et al. 2008).

Insert Figure 2 here

Examples

The scrutiny work shows that whilst overall, mathematics examinations tend to be devoid of context, there are included a substantial number of items that present mathematics as being a human activity, used by people in the world. The examples in this article are chosen explicitly to illustrate these; they are also relatively brief. They do not necessarily reflect the tenor of the examination paper from which they are drawn, and they are not necessarily typical of that examination as a whole.

Figure 3 is an assessment item in which a context is provided for required mathematical procedures, in this case, demonstrating skills of drawing a plan and an elevation, and estimating a standard volume. A problem solving element is also introduced, that of estimating the number of sheets on the roll.

Insert Figure 3

Whilst it may be possible to imagine circumstances when one might have to undertake this problem in life, actually that is not the purpose of the context, which is to provide an imaginative setting for a question to test the candidates understanding of plans and elevations, area and volume, and the relations between them. Thus the content domains were considered to be 50% measure (parts a and b), 25% algebra and 25% number (parts c and d). The item was considered to require largely procedural reasoning (parts a, b and c) with some analytical reasoning (part d) and a small sprinkling of interpretation (part d). The context for this question was categorised as 'artifical'. Note that the advice to scrutineers included the exhortation to use 'professional judgement' when making decisions and so this and other items generated qualitative discussions about meanings of the terms as well as the items themselves.

The following example (Figure 4), through implicitly indicating to candidates that mathematics may be useful and applied in real life, was considered to be mainly artificial with an authentic component. 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.

Insert Figure 4

On cursory inspection this question appears to offer an authentic application, but on closer examination we realise that the realism relates to specific mathematical considerations (Cooper and Harries 2002), 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. Readers as well as candidates may also object to the explicit stereotyped and gendered assumptions in the question which is clearly intended to appeal to girls. There is also an ambiguity, 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 5) to illustrate mathematics being used at work by a professional person. As the data looks as if it may have been invented the scrutiny recorded this context as mainly artificial. In this example the candidates are not required to forego other non-mathematical considerations, and so it can also be seen as routine calculation of a moving average, but with some interpretation and communication of the result suggesting to the candidate that there is a post-school world where mathematics is used, in this case the professional setting of accountancy.

In fact across all of the GCSE examinations scrutinised 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 two male 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.

Insert Figure 5

Discussion

Providing human settings in which mathematics may represent or be used to model real circumstances may be an important element of orienting assessment items towards the functional process skills of 'representing', 'analysing' and 'interpreting' mathematical situations. However, populating examination questions in itself is inadequate as a means of assessing expectations of functionality, that is, whether citizens are able to use mathematics in everyday life, including at work

and in educational settings. Indeed assessing mathematics under traditional examination conditions militates against this functionality.

Mathematics examination 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. The scrutiny 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. Questions frequently introduce a human element as if this is an important aspect of making functionality visible, whereas as we suggest above, the issue is more about the problem of modern tests being incongruous with real life.

Including people and animals in the assessment items gives mathematics examinations a friendly feel, although anthropomorphising mathematics can lead test questions into the realms of fantasy; with applications becoming spurious and unreal. The majority of mathematics examination items 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 (Noyes et al. 2010). Unfortunately simply representing mathematics as a human activity is not associated with functional process skills, rather the reverse as the context frequently serves only to liven up dull questions requiring routine answers.

We offer the beginnings of a typology of assessment items found in traditional GCSE examination that represent mathematics in context, viz:

- 'Real life', where assessments are set in the context of activities that the candidate might engage with, e.g. reading timetables, or scheduling tasks. However it is rare to find genuine data being used, and real life is mostly only used for the story line of the problem, as with the example of Michelle redecorating her bedroom. One danger of these story lines is that they easily become unacceptably stereotyped as well.

- Problems arising from genuine parameters (such as cost, time, dimension), but specifically mathematical in the skills, knowledge and understanding expected. 'Toilet Roll' is an example of this form. 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, often routine mathematical skills but located in in vocational settings, e.g. engineering, or as the moving averages example discussed, accountancy. Actually this is rarer than we assumed it would be, certainly in the scrutiny of 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 so as to present mathematics as something that will occur in future life, even though the skills being tested are securely located at school level.

- 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. Such problems as these rarely mention human beings at all.

What is in the mathematics assessment is closely connected to how the curriculum is experienced by students and teachers preparing for the examinations. Test items are used by teachers to prepare

students for future examinations, and so it is the assessment items that help them recognise the range of mathematical activity that is expected for successful performance to be achieved. The inclusion of a human element may make the curriculum seem more humorous and palatable for some students. It may even disguise for some the fact of the curriculum focusing on routines and not on problem solving. However, functionality is not simply a matter of including people into assessment items, as ironically this reinforces a routinised and procedural curriculum, rather than grappling with the challenges posed for incorporating genuine functional applications into school mathematics.

However, there are forms of assessment in mathematics other than examination under traditional test conditions that were developed during the Mathematics Pathways Project. For example, at both levels 1 and 3, information known as 'pre-release material' was provided to candidates and their teachers usually about three weeks in advance of the examination date. This material typically included sets of data or other contextual information to familiarise candidates with the contexts to be used before entering the examination hall. In theory, these pre-release materials provide the opportunity for teachers to help candidates navigate what might be asked, what is relevant and what is actually contextual window-dressing. The combination of pre-release material and unseen examination offers opportunity for exploring data and contextual problem posing. It also enables the use of authentic rather than contrived contexts, potentially extending opportunities to assess problem solving to include citizenship activities, such as managing personal finance, learning to save, and 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 offer potential for exploring mathematics as a modelling tool in realistic and relevant situations, predicated on real data and real information. 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, at level 3. These promote mathematics as a high level tool for solving problems, but there is no evidence of these settings at level 2 of standard assessments, so is is likely that currently students will meet this approach for the first time when (and if) they progress beyond level 2.

Importantly however for designers of assessment, pre-release materials raise questions over the technical validity and reliability of the assessment in the sense of measuring how well the assessment is matched to the intended learning outcomes, for the extent to which the examination is 'unseen' will probably vary from candidate to candidate, and from school to school. Just as the research team found when striving to reach agreement about the authenticity of items, the same is true for the context. As a consequence of cultural, economic or social factors affecting the ways that candidates engage with them, some contexts will be more familiar than others to some candidates. Also, in the lead up to examinations most schools and colleges are not working to timetable and so opportunities to spend time with students exploring the ideas raised in the pre-release materials are limited. Pre-release material is not currently used in mathematics at level 2 where the hegemony of 'unseen yet valid' examination dominates over 'seen and experienced' mathematics, firmly securing the dominance of the traditional mathematics examination for GCSE.

Yet the scrutiny suggests that assessment design at GCSE needs continued reform in order to address the functional mathematics agenda. A new secondary curriculum became statutory for 14-16 year olds in 2010 and has greater emphasis on problem-solving, functionality and mathematical thinking. This curriculum is already being reviewed for 2012. As part of the review, the fact of exemplification of the curriculum de facto through assessment items has been flagged as poor practice that needs to be addressed as 'the curriculum degrades into 'that which will be assessed' (Oates 2010, 16). It has been argued that the combination of pressure on schools to achieve highly in

GCSE, with the competition between awarding bodies for their share of the available candidates, drives standards down (NCETM 2011). We and others have argued elsewhere (Noyes et al 2010; Pope 2011) that pressure between awarding bodies to maintain their market share of schools has the effect of inhibiting change in examination as schools are judged by the success of their students. Those who are calling for standardised national assessment at GCSE may now, at last find a climate in which their voices may be heard.

Acknowledgements

[i] For full description of National Qualifications Framework please see
http://www.direct.gov.uk/en/EducationAndLearning/QualificationsExplained/DG_10039017
[ii] From August 2009 until April 2011 Qualifications and Curriculum Development Agency

References

Brown, M., P. Brown and T. Bibby. 2008. 'I would rather die': reasons given by 16-year-olds for
	not continuing their study of mathematics, *Research in Mathematics Education* 10 no.1: 3–18.

Burkhardt, H. 2007. Mathematical proficiency: What is important? How can it be measured?
	Assessing mathematical proficiency. Cambridge: Cambridge University Press.

Cooper, B. 2007. Dilemmas in designing problems in 'realistic' school mathematics: A sociological
	overview and some research findings. *Philosophy of Mathematics Education Journal* 20
	http://people.exeter.ac.uk/PErnest/pome20/index.htm (accessed 15 September 2011).

Cooper, B. and M. Dunne. 1998. Anyone for tennis? Social class differences in children's responses to national curriculum mathematics testing. *The Sociological Review* 46, no. 1: 115–48 .

Cooper, B. and T. Harries. 2002. Children's responses to contrasting `realistic' mathematics problems: Just how realistic are children ready to be? *Educational Studies in Mathematics* 49, no. 1: 1–23.

Gago, J.M. 2004 *Increasing human resources for science and technology in europe*. Brussels: European Commission.

Gainsburg, J. 2005. School mathematics in work and life: What we know and how we can learn more. *Technology in Society* 27, no. 1: 1–22.

Gates, P. 2002. Issues of equity in mathematics education: Defining the problem, seeking solutions. In *Teaching mathematics in secondary schools: A reader*. London Routledge Falmer.

Hodgen, J. and D. Wiliam. 2006. *Mathematics inside the black box: Assessment for learning in the mathematics classroom*.

Hodgen, J., D. Pepper, L. Sturman L., and G. Ruddock. *Is the UK an Outlier? An international comparison of upper secondary mathematics*, The Nuffield Foundation.

Joint Council for Qualifications. 2012. *Examinations results: A-levels* http://www.jcq.org.uk/national_results/alevels/ Accessed 7 February 2012.

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, no. 2: 235–251.

NCETM, 2011. National curriculum review final report: https://www.ncetm.org.uk/community/9046.

Noyes, A. 2009. Exploring social patterns of participation in university-entrance level mathematics in england. *Research in Mathematics Education* 11, no. 2: 167–183.

Noyes, A., P. Drake, G. Wake and R. Murphy. 2010. *Evaluating mathematics pathways.* Final

Report. Department of Education:

https://www.education.gov.uk/publications/eOrderingDownload/DFE-RR143.pdf

Noyes, A., G. Wake, P. Drake and R. Murphy. 2008. *Evaluating mathematics pathways.* Interim

report to the QCA stage 3. www.Nottingham.ac.uk/emp Available on request from the QCDA.

Noyes, A. and P. Sealey. 2012. Investigating participation in advanced level mathematics: A study of

student drop out. *Research Papers in Education*, 27, no.1: 123–138.

Oates, T. 2010 Could do better:Using international comparisons to refine the national curriculum in

england.

http://www.cambridgeassessment.org.uk/ca/digitalAssets/188853_Could_do_better_FINAL_in

c_foreword.pdf Accessed 12 April 2011

Pisa. 2009. *The OECD programme for international student assessment.*

http://www.pisa.oecd.org/dataoecd/51/27/37474503.pdf Accessed18 October 2010.

Pope, S. 2011. The English assessment regime: How consistency and standards stifle innovation and

improved validity for the assessment of mathematics. *Proceedings of the British Society for*

Research into Learning Mathematics 31, no. 2: 61–66.

Presmeg, N. and M. Van Den Heuvel-Panhuizen. 2003 'Leen Streefland's legacy' realistic

mathematics education research: Leen Streefland's work continues. *Educational Studies in*

Mathematics 54, no. 1: 1–4.

Rizvi, F. and B. Lingard. 2010. *Globalizing education policy*. London: Routledge.

Roberts, G. 2002 *Set for success. The supply of people with science, technology, engineering and*

mathematical skills. London: The Stationery Office.

Roper, T., J. Threlfall and J. Monaghan. 2006. Functional mathematics: What is it? *Research in*

Mathematics Education, 8: 89–98.

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. London: The Stationery Office.

Steen, L.A. ed. 2001. *Mathematics and democracy: The case for quantitative literacy.* Princeton, NJ: National Council on Education and the Disciplines.

http://www.scribd.com/doc/6074378/Mathematics-and-Democracy-The-Case-for-Quantitative-Literacy downloaded 16 September 2010.

Stobart, G. 2008. *Testing times: The uses and abuses of assessment*. London: Taylor and Francis.

Tomlinson, M. 2004. *14-19 Curriculum and qualifications reform*. Final Report of the Working Group on 14-19 Reform. http://www.dcsf.gov.uk/14-19/documents/Final%20Report.pdf Accessed 3 November 2010.

QCA. 2007. *Functional skills standards, 'introduction to mathematics'*. http://nationalstrategies.standards.dcsf.gov.uk/node/16057 Accessed 10 March 2011.

QCDA. 2009. *Functional skills criteria for mathematics Entry 1, Entry 2, Entry 3, level 1 and level 2* QCA/09/4267

http://www.qcda.gov.uk/resources/publication.aspx?id=d65d7c47-9b81-4bbc-8aba-d20c10ef7d49 Accessed 10 March 2011

Vorderman, C. 2011. *A world class mathematics education for all our young people.* Report for the Conservative Party http://www.conservatives.com/ Accessed 7 February 2012.

Wolf, A. 2011. *Review of vocational education–the Wolf Report*. http://www.education.gov.uk/publications/eOrderingDownload/The%20Wolf%20Report.pdf: Accessed 11 March 2011.

Table 1: EMP Scrutiny of Mathematics Assessments 2008-2009

level	2008		2009		Totals	
	Papers	Items	Papers	Items	Papers	Items
1	18	281	14	218	32	499
2	23	366	16	223	39	589
3	15	139	8	45	23	184
Total	56	786	38	486	94	1272

Figure 1: Scrutiny Protocol *(Extract completed for Toilet Roll, Figure 3)*

Left instructions			Label	Value		
Identify *percentage* of question allocated to different **content** areas. You can use the marks allocated by question and mark scheme as a guide but the distribution **should be** based on your professional judgement of what the question is expecting of candidates.			**Question/Task no.**	7		
	Structure		**total marks**	8		
			no of parts	3		
		Number of marks per part	**1**	2		
			2	2		
			3	2		
			4	2		
			5+			
			Max mark for a part	2		
Allocate *percentage* of marks to appropriate **mathematical processes**. Refer to NC, Key Skills Standards etc. for clarification. You can use the marks allocated by question and mark scheme as a guide but the distribution should be based on your professional judgement of what the question is expecting of candidates.	**Content**	**Domain**	**Number**	25		
			Algebra	25		
			Geometry			
			Measures	50		
			Statistics & Probability			
	Process skills		**Representing**			
			Analysing- reasoning	20		
			Analysing- procedural	70		
			Interpreting	10		
			Communicating			
Identify the main **context**(s) of the task / question allocating the *percentage* to categories ■ pure (within mathematics itself) ■ artificial (context that can be understood / imagined but lacking authenticity), ■ authentic (meaningful situations that students or adults might be expected to want to explore or problems they might wish to solve).		**Context**	**Pure**			
			Artificial	100		
			Authentic			
		Complexity	**1**	50		
			2	50		
			3			
			4			
Judge the **complexity** of the task / question using a 4-point scale: 1: straight forward / routine 2: 3: 4: complex requiring synthesis of information across a number of domains. Allocate *percentages* using professional judgement.		**Unfamiliarity**	**1**	75		
			2	25		
			3			
		Barriers	**Language**			
			Gender			
			Culture			
			Other			
Judge **unfamiliarity** of the task / question using a 3-point scale 1: typically met throughout learning programme 2: has some novel aspects 3: situation unlikely to have been met before Indicate judgement with *percentages* in the appropriate cells.		**Calculator**	**useful**	1		
			not useful			
			not allowed			
			Measuring equipment	1		
		Computer	**useful**			
			not useful			

Calculators: If calculators allowed judge whether useful or not. Make your judgement for the task overall and place "1" in one of these cells

Barriers: indicate with a "1" if the task / question might pose a barrier to candidates because of
■ language
■ gender
■ culture, or
■ other issues. (Indicate in "notes" the issues raised.)

Measuring equipment: indicate if use of ruler / protractor required in any part placing a "1" or "0" in this cell.

Computers: If allowed judge whether useful or not. Make your judgement for the task overall and place "1" in one of these cells

Figure 2: Sample 'task type' and 'process skills' profile of one Level 2 GCSE examination.
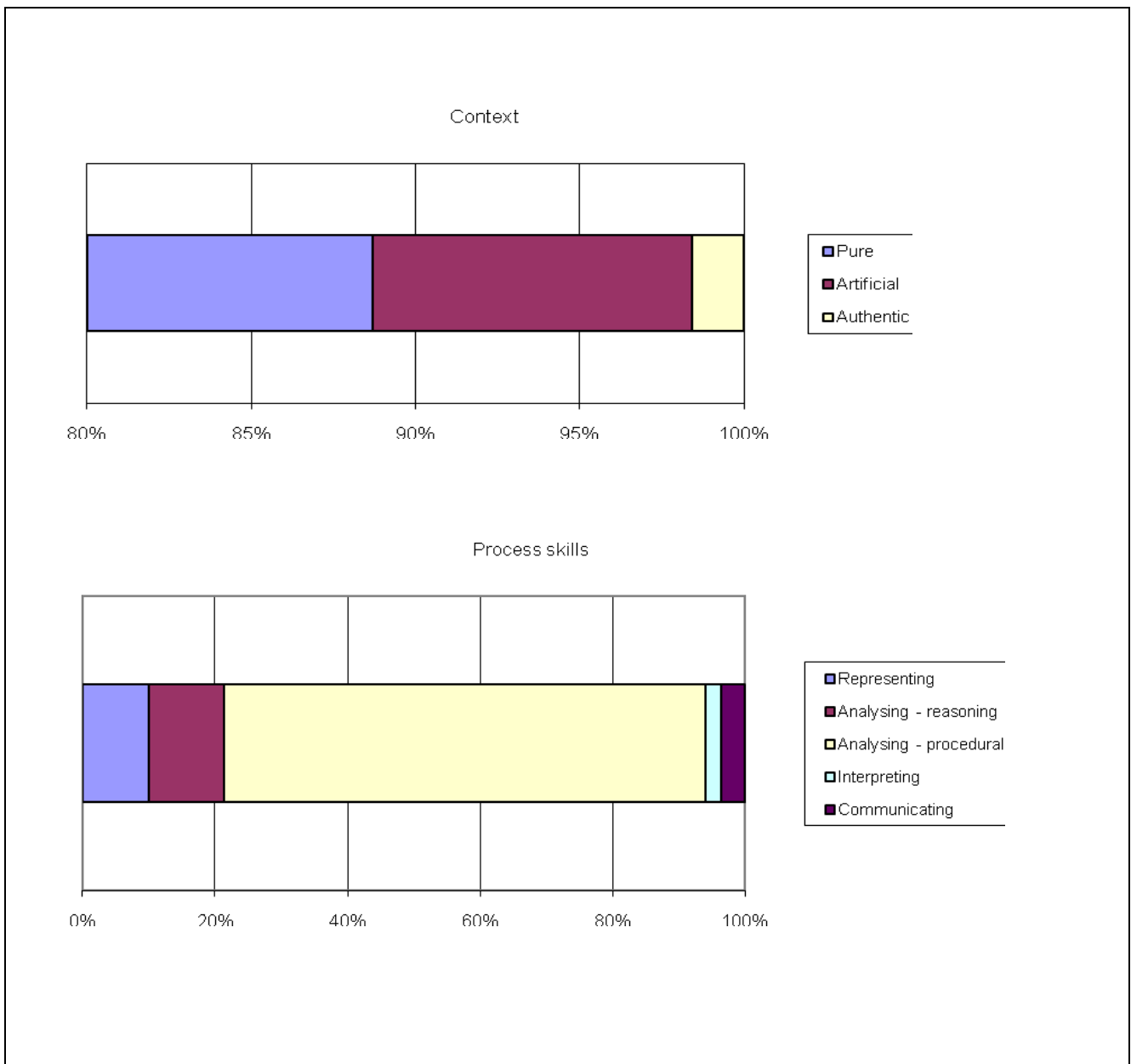
Figure 3

7 The diagram represents a toilet roll.

├──── 12 cm ────┤

├── 6 cm ──┤

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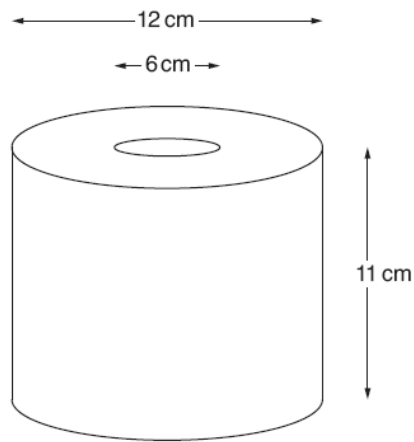
(a) Draw a full size accurate side elevation of the toilet roll. *2 marks*

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

(c) (i) By taking π equal to 3, find an estimate for the volume of paper on the roll. *2 marks*

(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

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

6 marks

Figure 5

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]