Context in Mathematics examination questions

Jackie Greatorex Research Division

Introduction

For at least two decades educationalists have debated whether Mathematics examination questions should be set in context. The advantages of context in Mathematics questions include promoting mathematical literacy, using mathematical tools and thinking to make sense of the world (Debba, 2011; du Feu, 2001). The disadvantages incorporate the difficulties in finding real-life contexts in which school Mathematics readily fits, and therefore examination questions can contain artificial contexts that require learners to make unrealistic assumptions (Clausen-May, 2006; Little and Jones, 2010). This is a contemporary debate. For instance, on 8 October 2013 the Government announced they were supporting schools and colleges to teach *Core Mathematics qualifications* (Department for Education, 2013). Solving significant problems in contexts is at the centre of the Core Mathematics qualifications. Guidance suggests that awarding organisations should assess Core Mathematics in context, using contexts suggested by higher education and industry (Browne *et al*., 2013).

The aim of this article is to revisit the debate to answer the following questions:

1. What are the advantages and disadvantages of examining Mathematics in context?
2. What are the features of a high quality context?

Initially several taxonomies (categories or classification systems) of context are reviewed and the research methods for evaluating the effects of context are considered. Subsequently, the advantages and disadvantages of using context in Mathematics examination questions are explored, focusing on research about public examinations in secondary school Mathematics in England. The literature is used to make recommendations about context in Mathematics questions.

Taxonomies of context

Several authors argue that the term *context* is particularly difficult to define for Mathematics (Berry *et al*., 1999; Little and Jones, 2007; Vappula and Clausen-May, 2006). Berry *et al*. (1999) believe that no definition can be found and consider that it is more useful to think about routine and non-routine questions. Routine questions are those to which learners are likely to respond with a prepared routine consisting of a small number of stages. Non-routine questions do not fit this description.

Whilst there is a dearth of definitions of context, taxonomies describing context abound. Some taxonomies are concerned with Mathematics for learners up to and including 16 years of age (Ahmed and Pollitt, 2007; McCusker, Nicholson, and Ridgway, 2010; Mevarech and Stern, 1997; Vappula and Clausen-May, 2006; Watanabe and Ischinger, 2009) whilst other taxonomies relate to learners aged 17 or 18 years (Debba, 2011; Little and Jones, 2007). The taxonomies are summarised below to further describe context and introduce terms.

Mevarech and Stern (1997) refer to *sparse* versus *real* contexts of questions about linear graphs. Although no clear definitions are provided, their work implies that real contexts are everyday contexts, and sparse contexts are more school-orientated contexts and potentially abstracted from everyday life.

Vappula and Clausen-May (2006) argue that contextualised questions are those that include stories about real-life events, a pictorial or a verbal model.

Watanabe and Ischinger (2009) categorise contexts according to content, including:

- *Personal contexts* – which are of direct personal relevance to learners
- *Educational and occupational contexts* – which include scenarios that learners might contend with while at school, including somewhat artificial problems, or problems that would be encountered in a workplace
- *Public contexts* – which are scenarios experienced in everyday life such as reading part of a newspaper
- *Scientific contexts* – which occur when a question is in a science context such as presenting experimental data.

Debba (2011) uses work by du Feu (2001) to derive question types, according to the characteristics of their context. The categories are:

- *Context-free* – tasks with no context, often simple equations and one-step arithmetic
- *Real contexts* – real problems which mention any named individual(s), institution(s), artefact(s), organism(s) or product(s). These contexts are chiefly statistical. Data are used and quoted and a source is acknowledged
- *Cleaned contexts* – real-life contexts are simplified such that the question is accessible to the learner and suitable for the time constraints of an examination
- *Parables* – fictitious contexts ascribed to an anonymous person/company/organism
- *Contrived contexts* – devised to fit a particular mathematical point, regardless of whether they are relevant to real life.

Some authors also classify questions according to the purpose of the context. Vappula and Clausen-May (2006) argue there are two purposes. The first is getting the story across (explaining the context), rather than supporting the learner with the Mathematics. The second is providing a model for the learner to think with. In this second case, the question and the context start to guide the learner towards a solution.

Watanabe and Ischinger (2009) also offer a classification system which relates to the purpose of the context:

- *Zero order* – the context is not needed to solve the problem
- *First order* – the context is relevant and needed to solve the problem and judge the correctness of the learner's answer
- *Second order* – the learner must engage with both the mathematical problem and its context to solve the problem. Furthermore, to judge the correctness of the learner's answer, the examiner must consider it within the context.

There are similarities between the taxonomies structured according to the purpose of the context. Zero order contexts (Watanabe and Ischinger, 2009) are similar to contexts intended to get the story across and not to support the Mathematics (Vappula and Clausen-May, 2006). First and second order contexts (Watanabe and Ischinger, 2009) are similar to contexts that provide a model for the learner to think with (Vappula and Clausen-May, 2006). A limitation of the taxonomies is that they generally omit marking with the exception of the taxonomy by Watanabe and Ischinger (2009).

The taxonomies summarised so far are descriptive or are structured according to the purpose of contexts. There are four further taxonomies which can be used to evaluate the quality of a context.

Ahmed and Pollitt (2007) classify questions according to whether they are *focused* or *unfocused*. A focused question addresses the aspects of the context that are most salient in real life for the learners. Unfocused questions do not address the aspects of the context that are most salient in real life for learners. A more focused context helps to activate relevant concepts, rather than interfering with comprehension and reasoning.

Little and Jones (2007), Little (2008) and Little (2010) refer to various characteristics of contexts:

- *Accessibility* – the familiarity and comprehensibility of the context, including the comprehensibility of the language and the clarity of the match between a context and a mathematical model
- *Realistic* – the fit of the mathematical model to the real world:
  a. *Natural contexts* – match reality
  b. *Synthetic contexts* – configure reality to match the Mathematics and can reify abstract mathematical ideas
- *Authentic* – the relevance and usefulness of the solution to the real-world context.

Wiliam (1997) reports a taxonomy of context, although his refers to context in the presentation of mathematical texts in general. He lists three types of context in Mathematics teaching:

- *Maths looking for somewhere to happen* – contexts that have little or nothing to do with the Mathematics being taught. The context is to justify the subject matter. This is very similar to synthetic contexts mentioned above
- *Realistic Mathematics* – contexts with a structure that maps to the mathematical structures being taught
- *Real problems* – contexts in which the key aim is solving a problem, and Mathematics may or may not be needed to find a solution.

Finally, McCusker *et al*. (2010) claim that most statistics questions from General Certificate of Secondary Education (GCSE) Statistics/Mathematics can normally be categorised within one of the following categories:

- *Real world but irrelevant or uninteresting* – contexts about the real world which are irrelevant or uninteresting to the majority of learners taking GCSE
- *Real world but not age appropriate* – contexts about the real world which are not age appropriate for the majority of learners taking GCSE. Boaler (1993b) and Debba (2011) also make the point that many contexts are about the adult world as experienced by examination setters, rather than the world as experienced by children
- *Context irrelevant* – contexts which are irrelevant to the real world
- *Unrealistic context* – contexts which present unrealistic information.

This implies that contexts in GCSE questions about statistics are generally flawed. Whilst there are several taxonomies of context, they are broadly similar. For the purpose of this review, context is taken to include all of the taxonomies discussed above.

Methods used to research the effects of context

Researchers use a variety of methods (often in combination) to investigate how context influences learners' behaviours, cognition and marks. The most prevalent design is a mixed methods study combining a quantitative experiment and a qualitative investigation. There are several experimental designs including matched pairs where learners in the control and experimental groups are matched for ability, attempt different versions of questions and all marks are quantitatively analysed. The methods for the qualitative investigation are wide ranging:

- An analysis of learners' responses to questions (Boaler, 1993a; Fisher-Hoch, *et al*. 1997; Vappula and Clausen-May, 2006)
- Interviewing learners about their experience of answering the questions, once they complete the questions (Ahmed and Pollitt, 2001; Crisp *et al*., 2008; Hong-Kim and Goetz, 1994)
- Stimulated recall sessions when learners are played a recording of themselves answering a question and asked to explain how they attempted the question (Ahmed and Pollitt, 2000)
- A questionnaire completed by learners about their experience of answering the question (Clausen-May, 2006; Khateeb, 2008; Little, 2010; Little and Jones, 2010; Song, 2011)
- An analysis of learners' written explanations of how they attempted questions (Cooper and Harries, 2002).

Mixed methods studies also combine quantitative analysis of marks with qualitative analysis of the questions (Watanabe and Ischinger, 2009). Other researchers add to this design by asking learners to complete questionnaires about their experiences (Cresswell and Vayssettes, 2006), or qualitatively analysing answers and interviewing learners (Debba, 2011).

There are fewer studies with a predominantly qualitative approach. Little (2008) and Morgan, Tang, and Sfard (2011) qualitatively analysed examination questions. Cooper (1998) qualitatively analysed responses to examination questions and interviewed learners.

Verbal protocols are a prevalent method of investigating the mental processes used to undertake a task (Kasper, 1998), including testing English as a second language (Green, 1998; Taylor, L., 2005). Verbal protocols are conspicuous by their absence from the list of qualitative methods above. In verbal protocol research, learners provide concurrent 'think aloud' verbal protocols as they answer an examination question (saying what they think as they tackle the problem) and retrospective verbal protocols (explaining their thought processes after completing a

task) (Page and Rahimi, 1995; Taylor, K. L. and Dionne, 2000). The verbal protocol data is qualitatively analysed to describe the learners' mental processes. Sometimes the combination of concurrent and retrospective verbal protocols is referred to as *cognitive laboratory* or *cognitive lab* (Levine *et al*., 2002). Cognitive labs are prevalent amongst major testing organisations in the USA (King and Laitusis, 2008; Test Measurement and Research Service, 2010a; Test Measurement and Research Services, 2010b; Zucker, Sassman, and Case, 2004). They are used to investigate a variety of domains: reading, Mathematics, language, spelling, listening, Science and Social Science (Test Measurement and Research Service, 2010a; Test Measurement and Research Services, 2010b; Zucker *et al*., 2004). This suggests that future research about context in GCSE, A levels or equivalent qualifications might benefit from using verbal protocols/cognitive lab.

Examining in context; advantages, disadvantages and recommendations for practice

There are several advantages and disadvantages to testing Mathematics in context, which are reviewed below. According to Brown (1999), examination questions should concentrate on the essence of the school subject, and context is best used in projects when learners have a teacher to guide them. Over a decade later this is still the prevailing expert wisdom. Therefore Browne *et al*. (2013) suggest 50 per cent of the assessment credit in the new core Mathematics qualifications is internal assessment. However, sometimes context is required in examinations. Therefore, suggestions for developing high quality contexts derived from research findings are presented below. They are presented with the caveat that the construct being tested influences what makes a high quality context. For instance, familiar contexts are more accessible to learners and are therefore arguably of high quality. However, if the construct to be tested is whether learners can apply mathematical principles in new situations, then using familiar contexts is invalid. Consequently the suggestions for developing high quality contexts should only be followed if they do not compromise validity.

The following section is organised by assessment topics, such as the construct.

Construct

GCSE Mathematics or A level Mathematics are designed to test school Mathematics as described in specifications, the National Curriculum, Ofqual's subject criteria and other guidance. The extent to which school Mathematics includes any topic or skill may vary over time. In other words, what counts as mathematical knowledge, skills and understanding varies. For the purposes of validity it is important that context does not detract from testing school Mathematics, for instance by encouraging responses using non-mathematical knowledge or creating construct-irrelevant difficulty (non-mathematical difficulty).

Unfortunately, these hazards are found in questions used in some research. Cooper (1998) found that some context questions elicit everyday knowledge or real-world experience rather than mathematical knowledge. Ahmed and Pollitt (2007) found that generally focused questions avoid construct-irrelevant difficulty. Furthermore, context can be successfully used to test whether learners can solve problems in new situations (Ahmed and Pollitt, 2007).

There are aspects of school Mathematics which cannot be tested in examinations with a time limit. For instance, contexts which are sufficiently open for learners to negotiate their own context and develop deep meaningful understandings are too lengthy for an examination, although they can be used for projects (Boaler, 1993b). Examinations cannot test aspects of the modelling cycle, such as discussing assumptions, and refining and critical reading of longer arguments (Little, 2010). Therefore, examination questions (with and without context) are restricted to requiring learners to undertake pseudo modelling. Real-world contexts in examination questions can be perceived as embryonic modelling problems (Little and Jones, 2010).

There are some areas of school Mathematics for which testing in context is clearly suitable. Context in data handling questions provides meaning for the questions; context-free data handling questions may reduce performances as the questions are devoid of meaning (Fisher-Hoch *et al*., 1997).

Therefore, context should:

- generally be used for all data handling questions
- be focused (address the aspects of the context that are most salient in real life for the learners).

Context

The weight of evidence suggests that context in itself is not the main factor making performance generally better or worse. That is context per se is not the main factor influencing the difficulty of questions. Difficulty is quantified in facility values (mean performance on a question, expressed as a percentage of the marks available for the question). Context does not hinder any more learners than do abstract presentations (Vappula and Clausen-May, 2006). Context in questions does not generally advantage learners who experience either of the following two approaches to teaching and learning (Boaler, 1998). One approach is conducting projects with considerable independence until just before the GCSE examination, when learners practise examination questions. The second approach is working through a textbook, asking for help with each new exercise and preparing for examinations. Context does not uniformly alter performance because each learner's individual experience influences how they interpret the context forming their own individual understanding of the context which impacts how they answer the question and how many marks each learner achieves (Boaler, 1993b; Debba, 2011).

If context *per se* is not generally influencing performance and difficulty, then what is? Learners' choices of Mathematics procedures are likely to be determined by the testing situation, rather than by the context (Boaler, 1993b). Competencies, content, item format (complex multiple choice, multiple choice, short answer, multiple short answer and extended response) and word count of an item are all better predictors of difficulty than context (Watanabe and Ischinger, 2009).

Nonetheless, context can shape learners' answers. Contextualisation in division questions encourages pupils to use informal or drawn methods (Vappula and Clausen-May, 2006). When encountering a context-free question part within a contextualised question, many learners fail to move their thinking out of the context and do not realise that the question part requires a simple textbook answer (Ahmed and Pollitt, 2000).

Therefore, context should be applied to all parts of the question.

So far this article has considered the influence of context *per se*. It is possible that context alone does not modify performance, but that aspects of context might shape the responses of individual learners or particular groups of learners. In the remainder of the article the influence that particular aspects of context have on learners' responses and performance are considered.

Realism

There are several advantages to incorporating real-world context in Mathematics examination questions. They reinforce the perception that Mathematics is useful (Little and Jones, 2010). Furthermore, real-world context can help ease the difficulty of questions by providing mental scaffolding for thinking within the context (Little, 2010) and helps learners use context-specific heuristic strategies (Little and Jones, 2010).

However, it is difficult to find real-life contexts in which school Mathematics readily fits (du Feu, 2001). Therefore, many Mathematics examination questions have synthetic contexts, which have disadvantages. Some learners perceive real-world contexts as artificial (Little and Jones, 2010). Context questions often require learners to make unrealistic assumptions (Clausen-May, 2006).

Despite these issues, synthetic contexts can be desirable as they are artificial models of a context which learners can be asked to evaluate. Learners can be asked to discuss and compare models and hence appreciate the relationship between reality and mathematical models. In such questions, synthetic contexts present limited mathematical models (Little, 2010).

Therefore, context should:

- be realistic or ask learners to evaluate synthetic contexts and the relationship between reality and mathematical models.

Relevance/familiarity

The weight of evidence suggests that if the context is unfamiliar or irrelevant to learners' lives then the demand of comprehending the question and the question difficulty increases. Familiar contexts have statistically significant positive effects on learners' inference making, and unfamiliar contexts can create disadvantages for deep comprehension (Song, 2011). Learners from rural poor areas tend to gain few marks on question papers with contexts which are irrelevant to their lives (Vurayai, 2012).

However, there is contradictory evidence on this issue, even from within the same research. For instance, Debba (2011) reports conflicting findings: unfamiliar contexts can be a barrier to performing well, but the level of real-life relevance that learners attribute to a context does not relate to performance.

Context triggers learners' schemas of everyday experience and reasoning which they use to answer questions, and can result in wrong answers (Ahmed and Pollitt, 2000; Debba, 2011). For instance, one learner's free ATM withdrawals influenced their answer to a question asking learners to work out an ATM cash withdrawal fee based on the amount of money withdrawn (Debba, 2011).

When contexts correspond to real-world situations with which learners are familiar, this can cause them to be unsure about whether to answer a question in terms of the subject or everyday knowledge and whether subject knowledge or everyday knowledge is rewarded with marks (Ahmed and Pollitt, 2000). Focused questions generally provoke fewer misunderstandings than unfocused questions (Ahmed and Pollitt, 2007).

Context can help learners answer a question when everyday reasoning and the correct method of answering the question coincided (Debba, 2011). For example, a football fan answered a question about teams and points in a league system. Their experience of the game and the point system helped identify the crucial information for solving the Mathematics in the question (Debba, 2011).

The challenge of writing questions with contexts that are familiar or relevant to learners is that they do not have a uniform experience, nor do they have the same experience as test developers. Therefore, some learners find a context relevant to their real lives, and others do not (Debba, 2011). Furthermore, many contexts represent adult life, or present adult metaphors, as they are written by adults (Boaler, 1993b; Debba, 2011; McCusker *et al*., 2010).

Cresswell and Vayssettes (2006) claim the following are typically encountered in the lives of 15 year olds:

- Tasks involving quantitative, spatial, probabilistic or other mathematical concepts. For example, media outlets (newspapers, magazines, television and the internet) are filled with information in the form of tables, charts and graphs about subjects such as weather, economics, medicine and sports
- Information on issues such as global warming and the greenhouse effect, population growth, oil slicks in the seas, or the disappearing countryside
- The need to read forms, interpret bus and train timetables, carry out transactions involving money and determine the best buy at the market.

McCusker *et al*. (2010), like Cresswell and Vayssettes (2006), advocate using information from the media. Furthermore, McCusker *et al*. (2010) suggest using contexts that are familiar and relevant to learners' lives, as well as context that are interesting to the learners. They suggest taking inspiration for context topics from a list of the most popular books from the learners' age range or their favourite televisions programmes.

In conclusion, context should:

- be focused
- be restricted to those for which the everyday reasoning and the mathematical reasoning used to answer the question are the same
- be relevant to learners' lives and familiar to learners.

Language

The language in questions can change their difficulty (Fisher-Hoch *et al*., 1997). The language describing contexts is sometimes a barrier to understanding the question requirements which can reduce performance, especially amongst learners with low levels of literacy (Debba, 2011). Furthermore, ambiguity in a question can lead to misinterpretations and learners giving the wrong answer (Cooper, 1998). Real-world contexts increase the word length of the questions, and if they are too long they can place too much emphasis on comprehension rather than Mathematics, thus reducing validity (Little, 2010). The word count of a question is a better predictor of difficulty than the type of context (Watanabe and Ischinger, 2009). Since GCSE was first examined in 1988, the readability of Mathematics examination questions has improved on general readability measures, that is the questions are easier to read (Morgan *et al*., 2011). This is due to the examination boards working to avoid obscuring the Mathematics with language (Morgan *et al*., 2011). This reduces the extent to which the Mathematics examinations measure

reading ability, and improves the extent to which they assess the intended construct (School Mathematics), thereby improving validity.

Learners expect questions to be written in positive language, and can accept the positive meaning of a statement as correct. Putting negative words in bold like **not** reduces, but does not overcome the problem (Crisp *et al*., 2008).

Therefore, context should:

- be easy to read or have an appropriate reading age
- use positive language
- use clear, concise, unambiguous language which does not obscure the question.

Resources

Resources are diagrams, graphs, illustrations and so on that are used to build a context. Pictorial analogues, rather than real-life contexts provide learners with powerful models to think with which enhance learners' responses (Vappula and Clausen-May, 2006). However, complex contexts, information-heavy contexts and diagrams containing a lot of irrelevant information can all lead to learners producing errors (Debba, 2011).

The resources in a question can prompt learners to use the wrong procedure or method (Boaler, 1993a). Learners generally expect that all resources are necessary to answer the question, and this influences responses to questions. For example, learners can place too much emphasis on a resource (Crisp *et al*., 2008).

Therefore, context should:

- provide pictorial analogues which are models to think with
- use only relevant resources and avoid unnecessary information.

Question layout

Question layout and the order in which information is provided can influence learners' responses. Question layout can shape learners' ability to find crucial information to answer the question (Debba, 2011). Inappropriate ordering of the problem and information within a question increase its difficulty (Fisher-Hoch *et al*., 1997). Question format can also prompt learners to use the wrong procedure or method (Boaler, 1993a).

Therefore, context should:

- present the problem and information in the order in which it is needed to answer the item
- clearly provide the information needed to answer the question.

Marking

The research is dominated by the issue of how context modifies learners' performance and responses. Unusually the taxonomy by Watanabe and Ischinger (2009) illustrates that context impinges on marking as well as answering examination questions. Marking open-ended contextualised questions reputedly requires a good deal of marker training and mark scheme development (Clausen-May, 2006). However, context questions can be designed to measure the intended knowledge, skills and understanding as well as be unchallenging to mark. This is achieved by listing possible solutions to a question, and asking learners to choose the correct solution or provide figures from the solutions (Clausen-May, 2006). Little (2010) reports a view that real-world contexts jeopardise reliability (test-retest and marking consistency). However, Little also found that well designed questions do not have these problems.

Therefore, context should:

- list possible solutions and ask learners to choose the correct solution or provide figures from the solutions.

Conclusions

This review describes taxonomies of context, explores the detail of how context modifies comprehension, performance and difficulty, and provides some suggestions for question writing. This conclusion focuses on wider issues.

Firstly, major testing organisations in the USA are committing a great deal of time and effort to cognitive laboratory and verbal protocol work. There appears to be less research about how learners comprehend and respond to GCSE and A level questions, and certainly less verbal protocol and cognitive laboratory work with learners regarding these examinations. This may reflect different examination development processes. In the USA, many examination questions are pre-tested, whereas in the UK GCSEs and A levels are not generally pre-tested. They are developed based on years of expertise instead. It is recommended that in future research considers verbal protocols and cognitive laboratories as methods of data collection for investigating the A level and GCSE question answering process, or how questions influence the qualitative aspects of learners' responses.

The review highlights that there have been improvements in contextualised questions in recent years. For instance, since the GCSE was introduced, the examination boards worked to avoid obscuring Mathematics with language (Morgan *et al*., 2011). Hence the questions are easier to read according to a number of readability measures (Morgan *et al*., 2011).

How context influences marking is under researched. Research in this area might find ways of improving marking in addition to those proposed by Clausen-May (2006). Furthermore, as new contexts are included in timed written examinations, such as contexts that Higher Education and employers suggest for core Mathematics, new ways of marking may be needed.

It is important to ensure that examination questions with context are of high quality. Several factors are better predictors of performance and difficulty than context. These include competencies, content, item format and word count (Watanabe and Ischinger, 2009). Therefore these factors, along with context, should be attended to in question development. Nonetheless, continued research to identify how context influences comprehension, performance and difficulty is valuable as it facilitates validity and guards against construct-irrelevant difficulty.

References

Ahmed, A., & Pollitt, A. (2000). *Observing context in action*. Paper presented at the International Association for Educational Assessment conference, Jerusalem, Israel.

Ahmed, A., & Pollitt, A. (2001). *Improving the validity of contextualised questions*. Paper presented at the British Educational Research Conference, Leeds. Available online at: http://www.cambridgeassessment.org.uk/Images/109682--improving-the-validity-of-contextualised-questions.pdf (Accessed 28 November 2013).

Ahmed, A., & Pollitt, A. (2007). Improving the quality of contextualized questions: An experimental investigation of focus. *Assessment in Education: Principles, Policy & Practice*, **14**, 2, 201–232.

Berry, J., Maull, W., Johnson, P., & Monaghan, J. (1999). *Routine questions and examination performance*. Paper presented at the Psychology of Mathematics Conference, Haifa, Israel.

Boaler, J. (1993a). Encouraging the transfer of 'school' mathematics to the 'real world' through the integration of process and content, context and culture. *Educational studies in mathematics*, **25**, 4, 341–373.

Boaler, J. (1993b). The Role of Contexts in the Mathematics Classroom: Do they Make Mathematics More" Real"? *For the learning of mathematics*, **13**, 2, 12–17.

Boaler, J. (1998). Alternative approaches to teaching, learning and assessing mathematics. *Evaluation and Program Planning*, **21**, 2, 129–141.

Brown, M. (1999). One mathematics for all? In C. Hoyles, C. Morgan and G. Woodhouse (Eds), *Rethinking the Mathematics Curriculum*, pp.78–90. London: Falmer Press.

Browne, R., Koenig, J., MacKay, N., Sheldon, N., Silcott, N., & Wake, G. (2013). *Report from the expert panel on core mathematics*. Available online at: https://www.acme-uk.org/media/13699/final%2007october2013,%20expert%20panel%20on%20core%20mathematics%20report.pdf (Accessed 28 November 2013).

Clausen-May, T. (2006). Reality and context in maths test questions. *Mathematics in School*, **35**, 5, 9–11.

Cooper, B. (1998). Assessing national curriculum mathematics in England: Exploring children's interpretation of key stage 2 tests in clinical interviews. *Educational studies in mathematics*, **35**, 1, 19–49.

Cooper, B., & Harries, T. (2002). Children's Responses to Contrasting Realistic Mathematics Problems: Just how realistic are children ready to be? *Educational studies in mathematics*, **49**, 1, 1–23.

Cresswell, J., & Vayssettes, S. (2006). Assessing Scientific, Reading and Mathematical Literacy. A framework for PISA.

Crisp, V., Sweiry, E., Ahmed, A., & Pollitt, A. (2008). Tales of the expected: the influence of students' expectations on question validity and implications for writing exam questions. *Educational Research*, **50**, 1, 95–115.

Debba, R. (2011). *An exploration of the strategies used by grade 12 mathematical literacy learners when answering mathematical literacy examination questions based on a variety of real-life contexts*. Master of Education, KwaZulu-Natal. Available online at: http://researchspace.ukzn.ac.za/jspui/bitstream/10413/5814/1/Debba_Rajan_2011.pdf (Accessed 28 November 2012).

Department for Education. (2013). *New maths qualifications to boost numbers studying maths to age 18*. Available online at: https://www.gov.uk/government/news/new-maths-qualifications-to-boost-numbers-studying-maths-to-age-18 (Accessed 11 October 2013).

du Feu, C. (2001). Naming and shaming. *Mathematics in School*, **30**, 3, 2–8.

Fisher-Hoch, H., Hughes, S., & Bramley, T. (1997). *What makes GCSE examination questions difficult? Outcomes of manipulating difficulty of GCSE questions*. Paper presented at the British Educational Research Association, University of York. Available online at: http://www.leeds.ac.uk/educol/documents/000000338.htm (Accessed 28 November 2013).

Green, A. (1998). *Verbal protocol analysis in language testing research: A handbook* (Vol. 5). Cambridge: Cambridge University Press.

Hong-Kim, Y., & Goetz, E. (1994). Context effects on word recognition and reading comprehension of poor and good readers: A test of the interactive-compensatory hypothesis. *Reading Research Quarterly*, **29**, 2, 178–188.

Kasper, G. (1998). Analysing verbal protocols. *Tesol Quarterly*, **32**, 2, 358–362.

Khateeb, M. (2008). *Cognitive Load Theory and Mathematics Education*. University of New South Wales Australia.

King, K. C., & Laitusis, C. C. (2008). *Sample Cognitive Interview Protocol*. Available online at: http://www.ets.org/Media/Research/pdf/Sample_Cognitive_Lab_Interview_Protocol.pdf (Accessed 28 November 2013).

Levine, R., Huberman, M., Buckner, K., & Goldstein, A. A. (2002). *The Measurement of Instructional Background Indicators: Cognitive Laboratory Investigations of the Responses of Fourth and Eighth Grade Students and Teachers to Questionnaire Items*. Available online at: http://nces.ed.gov/pubs2002/200206.pdf (Accessed 28 November 2013).

Little, C. (2008). *The functions and effects of real world contextual framing in A/AS pure mathematics questions: developing an evaluative framework*. Paper presented at the British Society for Research into Learning Mathematics, Kings College, London.

Little, C. (2010). *The use of real-world contextual framing in UK university entrance level mathematics examinations*. Doctor of Philosophy, University of Southampton, Southampton. Available online at: http://eprints.soton.ac.uk/166257/ (Accessed 28 November 2013).

Little, C., & Jones, K. (2007). Contexts for pure mathematics: an analysis of A level mathematics papers. *Proceedings of the British Society for Research into Learning Mathematics*, **27**, 1, 48–53.

Little, C., & Jones, K. (2010). *The effect of using real world contexts in post-16 mathematics questions*. Paper presented at the British Congress for Mathematics Education, Manchester University.

McCusker, S., Nicholson, J., & Ridgway, J. (2010). *Statistics assessment: The good, the bad and the ugly*. Paper presented at the Eighth International Conference on Teaching Statistics (ICOTS8), Ljubljana, Slovenia.

Mevarech, Z. R., & Stern, E. (1997). Interaction between knowledge and contexts on understanding abstract mathematical concepts. *Journal of Experimental Child Psychology*, **65**, 1, 68–95.

Morgan, C., Tang, S., & Sfard, A. (2011). *Grammatical structure and mathematical activity: comparing examination questions*. Paper presented at the British Society for Research into Learning Mathematics, Oxford University.

Page, C., & Rahimi, M. (1995). *Concurrent and retrospective verbal protocols in usability testing: Is there value added in collecting both?* Paper presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting.

Song, M. (2011). *Effects of Background Context and Signaling on Comprehension Recall and Cognitive Load: The Perspective of Cognitive Load Theory*. Doctor of Philosophy, Nebraska-Lincoln. Available online at: http://digitalcommons.unl.edu/cehsdiss/120/ DigitalCommons@University of Nebraska - Lincoln database. (Accessed 28 November 2013).

Taylor, K. L., & Dionne, J. P. (2000). Accessing problem-solving strategy knowledge: The complementary use of concurrent verbal protocols and retrospective debriefing. *Journal of Educational Psychology*, **92**, 3, 413.

Taylor, L. (2005). Using qualitative research methods in test development and validation. *Research Notes*, **21**, 2–4.

Test Measurement and Research Service. (2010a). *A cognitive lab report for the American Diploma Project Algebra I End of course exam*. Available online at: http://images.pearsonassessments.com/images/tmrs/tmrs_rg/AlgI_Cog_Lab_Report_101410_Final.pdf (Accessed 28 November 2013).

Test Measurement and Research Services. (2010b). *A cognitive lab report for the American Diploma Project Algebra II End of course exam*. Available online at: http://images.pearsonassessments.com/images/tmrs/tmrs_rg/AlgIICogLabReport_101410_Final.pdf (Accessed 28 November 2013).

Vappula, H., & Clausen-May, T. (2006). Context in maths test questions: does it make a difference? *Research in Mathematics Education*, **8**, 1, 99–115.

Vurayai, S. (2012). Poverty penalty in ordinary level mathematics in rural Zimbabwe. *Wudpecker Journal of Educational Research*, **1**, 5, 79–85.

Watanabe, R., & Ischinger, B. (2009). *Learning Mathematics for Life: A Perspective from PISA*. Organisation for Economic Co-operation and Development.

Wiliam, D. (1997). *Relevance as MacGuffin in Mathematics Education*. Paper presented at the British Educational Research Association Conference, University of York. Link to the paper as a Word document available online at: http://www.dylanwiliam.org/Dylan_Wiliams_website/Papers.html (Accessed 28 November 2013).

Zucker, S., Sassman, C., & Case, B. J. (2004). *Cognitive Labs* (Harcourt Technical Report): Harcourt Assessment, Inc.