Using a stimulated recall method to access test-takers’ cognitive processes when answering onscreen test questions.

Sarah Hughes
King’s College London and Edexcel

sarah.hughes@edexcel.org.uk


1 Background

It has been argued that the use of e-assessment in UK national tests is inevitable (Ripley 2004). The Qualification and Curriculum Authority for England (QCA) described its pilot ICT test as one of the main drivers for change in e-assessment and has identified national curriculum tests in mathematics and science as priorities for the future direction of e-assessment (Nesbitt 2006).

Edexcel began a suite of computer based testing projects to identify ways in which tests might be delivered onscreen. The maths strand of the project aimed to use the tools afforded by the technology to present and mark test questions which would be suitable for the assessment of national curriculum mathematics and which could not be delivered, tackled and marked on paper.

1.1 Research focus

The focus of the project as a whole is to establish what test takers are actually doing when they work onscreen, and is concerned with the validity of assessing mathematics onscreen. The focus of this paper is the development of a methodology, based on a stimulated recall method (Gass 2001, Calderhead 1981 and Lyle 2003) that will allow researchers to access evidence of test takers’ cognitive processes thus allowing research into the validity of such tests.

2 Theoretical framework
The validity of onscreen assessments is affected by a number of factors, but I consider these three to be of central importance:

1. The nature of the domain being assessed, in this case ‘school mathematics’. Added to the complexity of this domain is the changing nature of school mathematics given the impact of technology on teaching and learning (e.g. Tattersall, 2003).

2. The impact of the use of technology on the assessment. (e.g. Bennett, 2002)

3. The theoretical and practical understanding of test validity (e.g. Black 1998, Crooks 1996)

These three factors provide three interrelated bodies of theory and literature which have been used to provide a framework and to develop a method.

This paper presents an account of the development of, as well as a critique of, a method which aims to gauge the validity of an onscreen test by a process of accessing students’ cognitive processes and intentions when working on mathematics test questions.

3 Method

I carried out two phases of pilots. I describe and critique these in terms of how effectively they provided evidence of test takers’ cognitive processes. Two main pilots took place, which I shall call pilot 1 and pilot 2:

1. Pilot 1 - Stimulated recall using audio recording and observation notes to stimulate test-takers’ recall of their methods and strategies (after Lyle 2003).

2. Pilot 2 - Stimulated recall with video recording using digital video recording of the test-takers’ screen to stimulate their recall (after Calderhead 1995).

Alongside these two methods other considerations included whether to use individual or paired interviews (Arksey and Knight 1999) and if and how to incorporate test-takers’ own verbal reports on their thinking (Newell and Simon 1972, Ericsson and Simon 1980).
3.1 Pilot 1

In pilot 1 I attempted to stimulated test-takers’ recall using audio recording of their verbalisations as they worked through the test and my observation notes. In this section I describe the design of the interviews, the outcomes in terms of what evidence of cognitive processes was gleaned, and then reflect on the methods used and make suggestions for the next phase of data collection.

3.1.1 Background and design of pilot 1 interviews

In the first stage of the pilot I carried out four interviews, each with one ten year old. I started by using a verbal reporting method that I had used before to generate evidence of cognitive processes (Fisher-Hoch et al 1997). When attempting to determine the cognitive processes in learning and doing maths Gass (2001) recognised that as these cognitive processes were not observable there was a need to observe the behaviour resulting from these cognitive processes and use these resultant behaviours to probe learners further in order to attempt to identify cognitive processes.

One resultant behaviour of cognitive processes could be a verbal report generated during the task (concurrent) or following the task (retrospective). Newell and Simon (1984) argued that their verbal protocol method did not interfere with cognitive processes, although this was protested by Svenson (1989) who argued that where the task at hand was verbal the articulation of a verbal report did interfere because of the increased demands on short term memory. Svenson argued that there are three problems with verbal protocols: the limitations of individuals’ expressiveness, the length and complexity of a task and the difficulty of articulating automatic or tacit knowledge. The experiences I had supported the first two of Svenson’s criticisms: Previously (Fisher-Hoch et al 1997) I had tried to use a concurrent verbal protocol method (Ericsson and Simon 1980) and found this unproductive when working with children because of the demands of simultaneously verbalising whilst engaged in a task.

Gass (2001) described an alternative to verbal reports or think aloud procedures: stimulated recall. She described three factors in stimulated recall methods:

1) The support given to stimulate recall. Commonly this was a video recording of the task being carried out and the prompt to recall ‘what was going through your mind’.

2) The temporal relationship between the task and the recall. Gass described consecutive, delayed and non-recent recall.

3) The training of the interviewer and the interviewee.

Stimulated recall offered an alternative to verbal protocols that would not require pupils to concurrently verbalise and carry out a task. I chose to pilot a method which used retrospective verbal protocols with
stimulated recall. In the pilot interviews I collected and used a number of behaviours to attempt to stimulate pupils’ recall:

1) the computer screen which showed the final screen of the task they had just completed

2) my observation notes from watching the test-taker complete the task

3) test-takers’ notes or jottings, where they had made any.

Calderhead (1981) suggested that the cues provided by the audio tape or video will enable the participant to relive the episode to the extent of being able to ‘provide, in retrospect an accurate verbalised account of his original thought processes, provided that all the relevant ideas which inform an episode are accessible’. (Pg 212)

The process of interviewing comprised:

1. The test-taker is audio recorded as they complete the test. During the test the test-taker can use paper and pencil for jottings or notes. I took observation notes.

2. Immediately after completing the test, the test-taker and I discuss how they did the task using my observation notes and the final screen of each question as stimuli for discussion. This discussion is audio recorded. What I used as stimulated recall was simply my observation notes and the screen as it was left after the completion of the question.

3.1.2 Limitations of stimulated recall

Five limitations of stimulated recall have been identified in the literature. These are described below.

**Reflection interferes with recall.** Participants may report what they think on reflection and with hindsight, rather than reliving their strategies (Lyle 2003).

Omodei and McLennan (1994) compared stimulated response to free recall and argued that stimulated response results in more passionate, disorganised and illogical reports. Lyle proposed that pupils would react to what was viewed and not actually recall their thought processes. And so the resulting data was not about their cognitions, but a newly formed cognitions which contains reference and reflections on their initial cognitions. Yinger (1986) proposed that this ‘new view’ of the event is subject to the ‘luxury of meta-analysis and reflection’.

**Participants want to present themselves favourably.** McConnell (1985) warned of the danger that pupils will want to present themselves favourably and that this will impact on their verbalisations. When participants see the video or hear the audio and are asked to respond, what are they responding to? McConnell argued that this may be an exercise in self-criticism, rather than a recollection of their
behaviour.

**The task must be a goal-directed activity.** Hirst (1971) proposed that the stimulated recall method assumes that the action being researched (teaching in the case of Calderhead) is a goal directed activity. Test taking is a goal directed activity (although it can be argued that the test-takers is now always aware of the test-setters intended goal for them).

**Performance anxiety.** Fuller and Manning (1973) reported that confidence in performance or anxiety may influence this recall, or the extent to which the participant is prepared to report it. My pilot interviews suggest that it would be beneficial to use students who are confident rather than anxious.

**Are cognitive processes communicable in a verbal form?** In considering in what form mental representations of mathematical thinking might take it seems that they are likely to take other forms than just verbal, including spatial, visual, kinaesthetic and possibly auditory. The consideration of the communication of mathematics relates to the body of literature about mathematics as a language, and Language as a part of mathematics, its structures the processes. This relates to Vygotsky’s perception of language as a cultural tool used for developing shared knowledge within a community and for structuring the processes and content of thought within an individual mind (Mongahan 2005). Wergerif and Dawes (2004) stated that

> …it [maths] is also a kind of language. That is, maths can offer a form of social communication between people. To become fluent in that language, as with any language, children need guidance and opportunities to practice (p.102)

This makes me think that the school culture in which the pupils learn may influence how and what they are able and willing to articulate when working on tasks in the interview.

### 3.1.3 Outcomes and findings of pilot 1

In this section I report on how the data gathered from pilot 1a allowed me to answer the question: What cognitive processes are involved in answering the onscreen questions?

### 3.1.4 Analysis of pilot 1a

Pilot 1 method elicited only limited evidence of cognitive processes. A few interesting things that were found related to 1) test takers’ mathematical thinking and 2) my reflections on the method and implications for pilot 2. These are described below.

**Recall versus reflection** Hannah in pilot 1a (and Dan in pilot 1b also did this) described a strategy which was not the same as the strategy that I observed them using when answering the question.
The question was:

There are 7 ways to make two more squares red so that the black line is a line of symmetry.

Show them all.

Test-takers were able to view up to seven grids at one time on the screen.

Hannah described how she selected squares from a grid by moving around the grid horizontally along the top row, then descending down the last column. But my observation notes show that she did not follow this strategy: She used a rather less systematic approach in which she moved around the grid selecting squares that were next to each other.

This raises questions about why a test-taker’s descriptions of his or her behaviour may not match the actual behaviour.

- Was the process she used implicit?
- Was it forgotten?
- Was it communicable in verbal form?
- Was she attempting to present herself favourably?
- Was she suffering from performance anxiety?
- What impact was I, as her maths teacher, having on her thinking and reporting?
A key influence on reporting, I suggest, is that it reflects the learning that had taken place during the task and whilst reflecting on it. It was possible that she was not recalling her original method because she was still adapting her understanding and schema. Strategies are often post hoc (reference? This is a note from Jeremy), or at least start to emerge part way through answering or developing a concept of the question and what it means and requires.

Lyle (2003) warned of the risk of stimulated recall eliciting a reflection on them method used, rather than a straight, unmodified recall of the method used. Hannah could be describing to me, with the benefit of hindsight, a modified strategy which would be effectively used to answer the question she had tackled. And her description was subject to the benefit of meta-analysis and reflection. (Yinger 1986)

**Speed of response/answer**

There were a number of cases where the students got the question right, but I did not manage to capture evidence of how they got there.

Cognitive processes can be unconscious or maybe automatic or tacit (Polanyi 1967). Indeed, there was evidence of mathematical thinking in answers i.e. test-takers had given the correct answer and therefore had done something mathematical, but the method I employed to collect evidence of this mathematical thinking was not successful. It is not just a case of slowing down what is observed: seeing the observable actions of test-takers over every split second won’t help if the representations that I want to make explicit are not verbally communicable.

**Checking.** Hannah showed evidence of checking her answer by making a rough count of the number of squares to give her the area of a shape.

*Interviewer: Is that the first time you counted the area of the trapezium?*

*Hannah: Well, I had sort of roughly counted it*

*Interviewer: You’d estimated it?*

*Hannah: Yeah, but then I was just checking*

So what does that tell us about cognitive processes and mathematical thinking? Checking is a more easily distinguishable process, often the last step of a linear process, and as such easier to identify, hence it may be easier to find evidence of it than of more complex, parallel and interrelated processes.

**Test-takers’ experience of teaching and learning.** One of the three test-takers in pilot 1a used the jotting paper provided. This test-taker used it to record the horizontal calculations.
And when recalling, she described what she did.

Interviewer: So can you talk this through what you did next?

Luisa: 25 divided by 10 is 2.5, so 7 x 2.5 is…

Interviewer: yeah so 17.5 plus 7.5 is…

Luisa: 25.

This highlighted to me the importance of test takers’ experiences of teaching and learning. Luisa had recently moved to this UK primary school from her homeland of China. She was very confident with formal maths and calculations but not with communicating maths. I was not clear whether this was a manifestation of Luisa’s lack of English language skills more generally or a result of her experience of the teaching and learning of maths and resultant beliefs about what maths is. This also raised questions about her beliefs about the nature of school maths and how problem solving and informal methods might not be as valued by Luisa as students taught in the UK.

Context. Luisa said ‘I don’t really understand’ when presented with a question in the context of mixing paint. But nonetheless, and without help, she completed the questions and recorded the correct answer. Issues about Luisa’s understanding of culturally embedded maths, or maths in a UK everyday context aside, this raises questions about the authenticity and hence, validity of the question.

Language/instructions. The data did provide some limited information that the technology influenced their thinking. The three test takers in pilot 1a all, at some point, did not understand something.

For Hannah this seemed to be related to an instruction to make a rectangle, when actually there was not a rectangle visible on the screen, only two trapeziums.
Luisa was not clear how to ‘drag?’

And Lizzie showed that when working with an onscreen function machine she did not understand what the machine was doing behind the scenes. ‘I don’t know what it is going to do…sorry I’m a bit confused’.

In this case the computer was carrying out ‘unseen’ or hidden actions that were not shared with Lizzie, and this is very much unlike a paper activity, where the person is much more in control. Whereas Lizzie had to respond to and guess what the computer was doing. It all seemed a bit alien to her.

So although we see here three examples of being ‘stuck’ they are for slightly different reasons, but with an underlying similarity: the use of technology to present the question. I believe that the design and use of pictures/diagrams in Hannah’s question caused problems, and this interacted with the fact that the shape began as a trapezium and she had to act on it to change it to a rectangle, i.e. The dynamic nature of the question caused confusion. And only onscreen questions can be dynamic in this way. For Lizzie, again the non-static nature of the function machine she was using seemed novel and to require a different approach, or at least comprehension to build a schema for what the question required, was the sticking point. Luisa, has a specific problem with language in the question rubric which was related to how to respond to the question via the technology.

So the technology is the new aspect, but it interacts with age old accessibility problems in paper testing such as rubrics and wording.

3.1.5 Pilot 1b – Dan

After interviewing Hannah, Luisa and Lizzie and before interviewing Dan I heard Jim Ridgway speak. He talked about a method of assessing students which requires that students are asked what advice they would you to another student who was going to do a similar question. I saw the addition of this question to the interview method as a possible means of giving students an opportunity to reflect on their
answer all and put into action their meta-awareness. Hence the new framework that I tried comprised:

1. The test-taker is audio recorded as they complete the test. During the test the test-taker can use paper and pencil for jottings or notes. I took observation notes.

2. The test-taker and I discuss how they did the task using my observation notes and the final screen of each question as stimuli for discussion. This discussion is audio recorded. What I used as stimulated recall was simply my observation notes and the screen as it was left after the completion of the question.

3. The student describes what advice they would give another student who was going to do a similar task.

I also clarified and added instructions on how to verbalise and how you might think that despite some things seeming obvious, that they still should be said.

3.1.5.1 Analysis of pilot 1b

The data from pilot 1b was more effective than pilot 1a in terms of illuminating some cognitive processes. In some areas the findings reflected those of pilot 1a. These were the reading of instructions, the impact of context and the reflection versus recall issue.

Reading instructions. In pilot 1a the reading of instructions was raised because test-takers made mistakes interpreting question instructions. I proposed that these misinterpretations were because of the influence of technology on the test. Dan missed some key info on his first reading of a question which required that he fulfilled two conditions (shape and area): he only attended to one of the conditions, and on realising chastised himself, ‘I was careless, I didn’t look at that it had to be the same area’.

Context. One question was in the context of mixing paint, and asked test-takers to find out how much blue and yellow paint was required to mix 20 litres of green paint. This context, as Luisa found in pilot 1a, affected the process of answering: Dan said ‘you want twenty litres of green paint, which is quite a lot’ and showed an understanding of the context and how real it was (or was not). As a more able student Dan didn’t let that information distract him: was it a case of recognising the artificiality of a ‘real life’ context and disregarding it.

Reflection versus recall.

This excerpt of Dan’s interview showed evidence that he was reflecting on what he did – not recalling what he did. And his use of present tense suggests that he is still working on the question, rather than describing the process he went through to answer it in the past. It also seems that he is able to use meta-awareness (Yinger 1986), when he says ‘maybe a good tactic would be…’ this shows that indeed he
has not stopped working on the question and continued to improve or adapt his work.

Interviewer: Explain to me how you approached this question and what you thought it was all about.

....

Dan: OK, first two squares are red. If I was to make two or more squares red, so the line of symmetry, I would need... OK, it's one, two, from the line of symmetry, and that one's one, so I can put... again this is the same, and actually it's the same just tilting on it's side. Now I've got the diagonal, it's already symmetrical, but to make different grids we'd have to be using different squares every time, and we have to use two squares so we've got to put them in every square so once you know you've used all the squares maybe a good tactic would be... to fill in the squares to maybe go down the grids...

Interviewer: So you're going down the first column are you?

Dan: Yes, and so until you get to the bottom, which then you've got to fill in the ones in the middle, if you'd done it a different way you’d have filled in the ones in the middle and you get seven ways.

Another feature of the extract is my questioning, which I don’t think is clear enough in directing him to recall rather than reflect on his answering. Gass suggested that the prompt ‘What was going through your mind?’ was productive.

I don’t want to prevent test-takers from using their meta-awareness, but I need them to expand on it to be able to get them to describe how they did do it as well as how they would do it. There is an implication here for the questioning techniques used in interviews. I need to ask for reflection as well as recall. And in asking Dan to tell me what advice he would give someone else doing the question I was hoping to separate recall data from reflection data. However, what I did not do was make it explicit to him that was the two different questions wanted. I am reminded of Gass’s advice that both interviewer and interviewee should be trained in the technique and think that I need to be more explicit about the difference between recall and reflection and so allow test-takers to describe both to me.

Findings/evidence of other cog processes/mathematical thinking

The data from Dan’s interview was richer and gave more access to descriptions of cognitive processes, or at least descriptions of behaviours resulting from cognitive processes, than the data from pilot 1a. It was evident that his descriptions of how he would advise a peer tackling the question helped illuminate his cognitive processes. Nonetheless I need to remain aware that this would be because Dan is especially articulate and has good meta-cognition.

I saw Dan use processes including recall of knowledge and the use of physical cues. Then there were specific approaches to tackle individual questions and he was able to describe behaviour that resulted from his thought processes. Some of these are described below.
**Working backwards.** When Dan was describing how a peer could tackle a question he used a strategy that involved working backwards. He started with how much paint you need in the end (unlike the order of the information given). I looked for a way of reaching the target amount of paint with the proportions given.

Interviewer: Now imagine another student is about to do a similar type of question – how would you tell them to go about it?

Dan: I would see how many litres of paint they need…

Interviewer: altogether at the end?

Dan: Yeah. And then I would check how much it needs of one compared to the other, and maybe add them together to make 10. But if they wanted to make 15, you would have to have 7 plus half of 7. So basically I would look at how I could use the numbers equally doing the same operation to make the final number, and then I would find out what percent that would be of the number.

Watson and Mason (1998) described this type of ‘reversing’ as a mental activity typical of mathematical thinking. It looks like the new questioning technique employed in pilot 1b helped extract this type of thinking.

**Recognised a mistake.** The new questioning technique, of describing to a peer, allowed Dan to correct himself when advising a peer to ensure they wouldn’t make the same mistake that he made.

**Efficiency – choice of strategies.** When describing the advice he would give to a peer Dan recognised two ways to do a task; one by continuing a developing pattern and the other by calculating.

Interviewer: How would you explain to somebody how to do it with multiples of 8?

Dan: With the 8s, on a number grid like this it would go down one and left two.

Interviewer: Yeah…

Dan: Not all numbers would do this but, three 8s is 24, you could go down another and left another 2. So this would work with 8s, let’s say. 32… so the next one would be 50… so you have to work backwards…

…

Dan: yes, there’s a pattern you’ll notice with each of these. But in some cases it might be quicker just to work it out, let’s say 5s… 4s and 2s are pretty easy. Basically, I work up to the first three by adding and then I check for a pattern and if there was a clear pattern I could go point it out.

This also raised the question of what he means by ‘some cases’ and how I can ensure that I don’t miss the opportunity to pick up on things like that during the interview.
Re-using a previous strategy. He recognised similar questions and used a previous strategy again. Two other interviewees did that too, but were not explicit about it. This has implications for my questioning technique, and I need to ensure that I have a means of capturing this type of data.

Recognising similarity. When tilting a shape on a line of symmetry Dan could recognise similar shapes in different rotations and was explicit in describing these.

Recall of facts. There were times, as there was with the interviews in pilot 1a, that the test-taker recalled information. For example Dan said when questioned how he answered a question that ‘I just knew it’. Luisa and Lizzie both also described the recall of information in this way. The difference between Dan and the two girls was that they were unable to expand when probed, whereas Dan was able to differentiate the information that he recalled and what he did with it once recalled.

Interviewer: Well it’s a very effective strategy up to 90, and you got 99. Did you use the strategy to go from 90 to 99?

Dan: No, I just knew it. You can see there’s an obvious pattern, it goes down diagonally on the board.

And later on another question he showed recall of the knowledge that two parts of a whole constitute 50% each:

Dan: …since one part of blue and one part of yellow makes 2 parts of green, so they’re 50% each.

The implication is that I need to find a way to describe what was done as well as what was recalled. Maybe being explicit about this distinction with the test takers and asking them to describe both would elicit more data. However I am aware that my beliefs about mathematics, particular the value I place on processes in maths over recall of information, may skew the data collection to include more talk of processes.

3.1.5.2 Reflections on pilot 1b

- Dan is very articulate. I must consider what effect this had on the resulting data and how I can consider that in future data collection.
- The pilot 2 method took more time but was well worth it for the increased quantity and quality of data.
- I will consider extending the framework to also add a fourth aspect to the interview method which asks them to devise a similar question.
- Clarity of the data would be much improved by a video/replay of the test-taker’s onscreen actions.
3.1.6 Implications for pilot 2

Implications for pilot 2 are discussed below under the headings:

1. Use of video,
2. Questioning techniques,
3. Use of paired interviews,
4. Recording and formatting of data.

3.1.6.1 Use of video

At many points in the pilot interviews my data was lacking because I did not have a record of the pupils’ work. Not only would a recoding of the pupils onscreen work capture more data, but it could be used as a stimulus for their recall, and a prompt for them to remember what they were doing and why.

The use of video as a stimulant for recall could also help to reduce the tendency for test-takers to reflect upon and adapt their strategies, and just recall what they did when they first encountered the questions.

3.1.6.2 Questioning techniques

The introduction of the prompt to describe how to advise a peer doing a similar question created much more useful data. This could be expanded in pilot 2 to include the instruction to devise a similar question, and to discuss how the onscreen question could be presented and tackled on paper.

Also I think there is a need to be structured in the prompts used to stimulate recall, as I had tended in pilot 1 to adlib and, at points, provide feedback and slip in to a 'teacher role'.

3.1.6.3 Use of paired interviews

When working on the tasks, even although pupils were asked to articulate their thoughts as much as possible, they did make very few utterances. One way of helping to generate more verbal output whilst working on the questions could be to have pupils working in pairs. Arksey and Knight (1999) advocate this method because interviewees may fill gaps for each other and because it is easier to establish an ‘atmosphere of confidence’.

Evens and Houssart (2007) warn that interviewing in groups or pairs changes the role of the interviewer to that of a facilitator or moderator. So the introduction of paired interviews brings other considerations as well as solving problems identified in the pilot interviews.
I will also need to consider how to pair students to facilitate this verbal interaction. Fujii (2003) paired children with different answers to mathematical questions in order to introduce some cognitive conflict. However, Evens and Houssart (2007) found that interviewing a pair of children with conflicting answers did not necessarily result in more discussion between them: Some would not be concerned that their peer had a different answer, some would justify their own (incorrect) answer and others were not influenced by the conflicting answers. Lampert (1990) argues that a ‘culture of disagreement’ in a classroom can have a positive impact on mathematics learning and encourage a culture of explanation and justification. If I was to attempt to generate discussion through this kind of conflict then it will be necessary for me to identify if the students I interview are used to this kind of culture. I would argue, if this kind of conflict is unfamiliar then it could result in a negative experience for the pupils and result in unease. It will only be productive to encourage a ‘culture of disagreement’ where pupils’ response might be to expand on their thinking, rather than shut them down and shut them up.

A further possible impact of pairing children is the effect of interaction, collaboration or turn taking on their thinking. The context of my research is testing, which currently happens on individual students working alone. I must consider the effect that paired interviews might have on the validity of my research if I want my conclusions to inform a system of individual testing.

3.1.6.4 Recording and formatting of data

I need to reflect on how to format the transcript: there will be a number of different data types to present including my observation notes, the video, the discourse between test-takers during the test, the test-takers’ recall of their work.

Gass (2001) used this format for stimulated recall:

<table>
<thead>
<tr>
<th>Original episode</th>
<th>Stimulated recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Video</td>
<td>Pupil 1</td>
</tr>
<tr>
<td></td>
<td>Interviewer prompts</td>
</tr>
</tbody>
</table>

I adapted this format to allow for the inclusion of observation notes and both students’ responses.

<table>
<thead>
<tr>
<th>Original episode</th>
<th>Stimulated recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Video of screen</td>
<td>Description of pupil behaviour</td>
</tr>
<tr>
<td></td>
<td>Interviewer prompts</td>
</tr>
</tbody>
</table>
Previously I’d tried the format below, which was effective in presenting data for an individual interviews, but did not include my observation notes.

| Live conversation | Reflection | screen shots from the video of the screen – reconstructed by me from my observation notes |

3.2 Pilot 2

3.2.1 Background and design of pilot 2

In response to pilot 1 the interview for pilot 2 followed the process described in the instructions for interviewers below:

**Before the interview**
1. Interview children in pairs
2. Teachers can choose how to pair up children, but they need to be able to work together and be reasonably articulate.
3. Start the test.
4. Tell the children
   - that we are testing the question, and that we want them to try the questions, and then we will ask them about the questions.
   - That they will be audio recorded, to help us remember what they say
   - That the computer will be recording what they do and that we will look at that together after, so help them remember what they did.
   - They can ask a question at any time
   - That they will do two things:
     a. Do the six questions
     b. Watch the video recording and talk to us about what they did
5. Before they start the test make sure that Camtasia (screen recording software) and the digital voice recorder is operating.

**The interview**

**Stage 1 – the pair of students do the questions together**
- Make sure they taken turns controlling the mouse
- Make a note of anything they did or said that you’d like to follow up
- On completion Stop Camtasia recording and save the recording

**Stage 2 – talk through the video**
- Play the Camtasia recording and explain that it is a recording of what they did
- Use prompts below to try and find out why they made the decisions/took the actions that they did

**Prompts to get to their thinking:**
- What were you doing then?
- Why did you click/drag/move that?
What were you thinking then?
How did you know where to put your answer?
Did you agree on that?
What does …(symbol on screen)….. mean?
You seemed to be taking along time here, why?
Is there another way you could have done it?
What made that question so easy/hard?

After recalling for each question ask:

1. What advice would you give to someone else of your age if they were going to do this question?
2. How could you change that questions to make another similar question?
3. How would it have been different if that question was on paper?

Stage 3 - ask them for their comments
Allow them to comment or let off steam about anything.

Thank them for their help.

3.2.2 Pilot 2 findings

The data generated by the methods used in pilot 2 were more in quantity and quality than pilot 1. There was more discourse taking place during the test-taking, probably as a result of the paired interview. And there was more discourse generated when reflecting on taking the test. Probably as a result of the use of the video as a stimulus for recall. And of course, there were more questions from the interviewer with the addition of the request to describe how the test-taker would advise a peer to tackle the question.

Analysis of pilot 2 showed more evidence of cognitive processes, and more description of strategies and methods than in the previous pilots. When coding the data the following themes emerged, some relevant to cognitive processes, but more of them on related and wider issues.

1. Features of the question
2. Response types and behaviours
3. Learning / classroom experiences
4. Affect /opinion
5. Social dimension
6. Using technology
7. Strategy – identifying patterns
8. Strategy – other
9. Mathematical topics

An example of the data relating to a strategy used is shown below, in which Owen describes how he
visualises the pattern that multiples of eight made on a 1-100 number grid.

Sarah: OK. So that is the first one. You told me how you did it, you told me what advice you would give to someone else, and my third question about this is if you had to make up another question which was the same but used different numbers, what could you, what question could you make up?

Daisy: You could, say if you did something like the eight times table then you'd have to find a different pattern.

Sarah: Yes

Owen: You would, there is a different pattern. You could do the eight times pattern, times table, which would be slightly harder, quite a lot harder, its like a knight’s move.

Daisy: It would be like down and two left

Owen: Yes, one down two left. Kind of like a knight's move in chess. So you could do eight times or four times.

Probably prompted by the presentation of the number grid, Owen and Daisy described a visual representation of the multiples of eight. The method used allowed me to access a visual representation through verbal data, albeit on a question presented visually.

The extract above shows a strategy that could be used, not one that was used and so has given interviewees the opportunity to reflection, and even adapt their understanding of and approach to the question.

**Value of video**

The recording of the test-taker’s screen allowed me to more clearly see their actions. In Pilot 1 test-takers’ references to ‘there and there’ or ‘that’ could easily be lost, but with the video I could see at what the mouse was pointing or clicking. In this following extract Daisy is controlling the mouse and Owen is directing her.

Owen: Yes they need to be in the same place, relative to each other, say, there and there would work.

Daisy: There and there? No, no, no, we want like, those two, and then we could have those two.

Owen: OK, well we need, we only can add two, so I think there and there would work.

Daisy: But we've already just done that. (3 10)

Owen: No, there and there.

Daisy: There?
Using the video to stimulate recall allowed me to see what was meant by ‘there’.

**The visual nature of the screen**

The questions themselves tend to have a visual aspect, and the use of technology seems to be supporting the visual strategies. When writing the items it was apparent that creating shape and space tasks came more easily than number tasks. Below, Owen and Daisy described a visual strategy that they used.

Sarah: OK, right so um. So on this one where the line of symmetry is horizontal how did you know where to put the red squares?

Daisy: Ummm

Owen: At first I just imagined it was, that is how I imagined it, they are just two different things completely.

Sarah: The top and the bottom? Above and below the line of symmetry do you mean?

Owen: Yes. Above and below. Completely different. I put the (?) on the sides to try and make them look, almost the same really. It looks, coz if you think about symmetry you can make it look pretty much the same in your mind.

Daisy: It’s a bit like um, folding it in two making the same thing.

Sarah: Um (in agreement)

Daisy: Lets say you painted those two red squares and you just folded up the paper you’ve got them in the right positions.

Owen: Like you do, like you do in um, reception with the butterflies.

This shows that using a verbal interaction with me, the researcher, Owen and Daisy were able to communicate their visual and tactile representation of the problem. Implications for the next phase of data collection are presented in the following section.

### 4 Implications for data collection methods

The pilots raised some valuable questions and issues for planning the next phase of data collection which are discussed under the six headings
1. Reflection versus recall

2. Paper versus onscreen

3. Interviewee and interviewer training

4. The nature of mathematical thinking

5. The role of the researcher

6. Participants’ experiences of teaching and learning

4.1 Reflection versus recall

The measures implemented in pilot 2 encouraged both recall and reflection, i.e. questioning techniques to prompt both were used. There may be value in being more explicit in differentiating the two: this could be arrived at through the training of interviewees. As I reflect on the use of questioning techniques to differentiate test-takers’ recall and reflection I began to consider the order in which prompts were given. If learners are still adapting their understanding it may be more productive to allow them to complete this proves before asking them to recall their initial strategy. I am considering ordering the question to allow them to do this.

4.2 Paper versus onscreen

The three areas which come together to provide a theoretical framework for this research are:

- Test validity
- The nature of school maths
- E-assessment.

In order to evaluate the validity of an e-assessment of school maths it seems necessary to have as a benchmark or comparison a paper version of the questions. This would allow me to assess the impact of e-assessment on validity.

How might this be done? The feature of the electronic questions was that they were not replicable on paper, i.e. that they used the ‘affordances’ (Kennewell) of the technology. A valuable activity in itself, as a means of investigating validity of the questions designed to assess school maths, would be to develop comparable paper based questions. As many of the onscreen questions involve animations or dynamic visuals, which would be difficult to reproduce on paper. It may be possible to develop a version of the question that uses manipulatives alongside paper questions. The aim would be to then assess the
processes and understanding students use when tackling these different versions of question and compare them.

4.3 Interviewee and interviewer training

Gass (2001) suggested that in order for stimulated recall to be successful both interviewer and interviewee should be trained in the techniques of the method. The pilots have demonstrated that cognitive processes can be tacit, easily forgotten and that the recall of them is subject to interference of learning and meta-awareness. Training could be tried as a means of overcoming these problems. Of particular value, I think, would be to train test takers in the difference between recall and reflection, to allow them to provide both.

4.4 The nature of mathematical thinking

My aim is to access the cognitive processes used in mathematical thinking. I need to further understand the forms of mental representations that would be used when thinking mathematically, and know how researchers have identified these or made them explicit.

In the pilot analysis I became aware that linear sequences of behaviour were easier to identify and describe than parallel or interrelated behaviours and actions: the data collection and analysis methods will need to capture both of these types of processes.

4.5 The role of the researcher

The test-takers interviewed in the pilots had all been students of mine. I think that this had impacted on their behaviour, at least in terms of knowing what I value in student responses. In the next phase of data collection I will work with students unfamiliar with me, other than as researcher. As a note to myself, I need to begin data collection with an awareness of what relationship and role I want to develop between me and the students and schools. This would also be impacted if I chose to train interviewees because the interviewee/researcher relationship would be more intense.

4.6 Participants’ experiences of teaching and learning

Relating to the nature school maths is the test-takers’ experiences of teaching and learning. In the pilot I experienced students’ different approaches to maths with varying emphasis and uses of communication, formal methods and informal jottings. I believe that how students interact with assessment questions will, in part, be mediated by their experiences of learning maths and the messages they have received about the nature of maths and the value of its different components.

There are implications of this are for 1) my sampling technique and 2) the need to observe and understand
the test-takers’ experience of learning and maths and account for these explicitly in my analysis.

5 The next phase of data collection

The focus of the next phase of data collection will continue to be the validity of e-assessment of school mathematics. The three areas related being test validity, the nature of school maths and e-assessment. My aim is to identify what students are doing when they tackle mathematics tasks on computer and on paper to compare the mathematical thinking involved in both of these and then gauge if the mathematical thinking which results from these tasks can validly be considered ‘school mathematics’.

Features of the next phase of data collection need to include:

- Paired interviews with test-takers
- The use of stimulated recall, using screen recording, audio recording, my observation notes and students jottings as stimuli.
- The comparison of paper and onscreen (and possibly a version using manipulative objects) versions of similar tasks.
- The use of the four stage interview including
  1. Stimulated recall using a video of the test-takers’ screen as a stimulus.
  2. What advice would you give peer doing this question?
  3. Could you devise a similar question?
  4. How would this question be different if it were on paper?
- Observation/data collection on the students’ experience of teaching and learning maths.
References


23


*This document was added to the Education-line database on 30 January 2009*