University of Cambridge Local Examinations Syndicate

On-line assessment: the impact of mode on students' strategies, perceptions and behaviours

Martin Johnson and Sylvia Green

A paper to be presented at the British Educational Research Association Annual Conference, Manchester, September 2004.

Disclaimer

The opinions expressed in this paper are those of the authors and are not to be taken as the opinions of the University of Cambridge Local Examinations Syndicate (UCLES) or any of its subsidiaries.

Contact details
Martin Johnson and Sylvia Green
Assessment Directorate
University of Cambridge Local Examinations Syndicate
1 Hills Road
Cambridge
CB1 2EU
Tel: 01223 553843
Email: johnson.m@ucles.org.uk
green.s@ucles.org.uk

On-line assessment: the impact of mode on students' strategies, perceptions and behaviours

Introduction:

Modern technology offers a number of opportunities for education and assessment. Such opportunities also raise a number of important questions which research may help to address. The advantages of using computer technology for educational assessment in a global sense have been recognised by various commentators (Bennett, 2003; Raikes & Harding, 2003; Greenwood, Cole, McBride, Morrison, Cowan, & Lee, 2000) and these include lower administrative cost, increased adaptation to individual student characteristics and less demand upon teachers. Whilst recognising these system-level advantages it is important to explore the relationship between assessment mode and the behaviour of the students being assessed. In the literacy field a debate has recently developed concerning the effects of mode of communication on thinking, and this debate may have implications beyond the confines of literacy. Bearne & Kress (2001) use the term *affordances* to describe 'what is made possible and facilitated, and what is made difficult or inhibited' by a medium (2001, p.91). It is possible that the affordances offered by computer mediated assessment may affect the perceptions of students involved in computer-based assessment differently than if they were engaged in paper-based assessment.

Aim:

The transition from paper-based to computer-based assessment raises a number of important issues about how mode affects performance. This study follows on from a study that investigated the differences in performance of 10 and 11-year-old children attempting matched mathematics questions on-line and on paper (Johnson & Green, 2004). Attempting to understand why performances differed, this study explores the effect that mode had on the perceptions and behaviours of children taking part in the initial project.

A number of studies have already identified a relationship between assessment mode and student perception but, as Bennett (2003) points out, few have investigated this relationship with children of primary or elementary school age. Where this has been done it has been found that questions presented on computer were generally harder than when presented on paper (Choi & Tinkler, 2002; Coon, McLeod & Thissen, 2002). In a study with secondary school

students Greenwood et al (2000) found that any medium effect was most pronounced when questions involved spatial awareness or gross motor skills, with these questions being harder on computer than paper.

Other studies have attempted to investigate factors that influence children's attitudes to assessment mode. Richardson, Baird, Ridgway, Ripley, Shorrocks-Taylor & Swan (2002) worked with more able 9 and 13-year-olds and found a generally more favourable reaction to answering questions on computer compared with paper. There were a variety of reasons given for preferences. Most students alluded to having a generally positive attitude to computers and this affected their stance towards answering computer-based questions. This hints at the possibility that student motivation for computerised tests may be influenced by their experience of the medium beyond an educational context, and that this attitude is different for paper-based tests. This reinforces work done by Levin & Gordon (1989) who suggested that the dominant consideration affecting student attitudes to working on computers was their prior level of computer experience. This may also help to explain a more contemporary finding by Bridgeman, Lennon & Jackenthal (2002) who reported the comparative popularity of computer-based tests over paper-based tests in a study with American high-school students.

Richardson et al (2002) found other reasons that contributed to student preferences for computer-based questions were related to concrete differences between the questions in the different modes, such as the use of colour illustrations. Other reasons were more clearly related to the affordances allowed by the medium. Some children preferred computer questions because they involved less writing and because revised answers were neater than if they had been erased on paper.

This study seeks to explore whether Choi & Tinkler's finding that computer-based questions were more difficult translate from an American elementary school context into an English setting. Further, it aims to investigate whether the concerns which Richardson et al found to affect the attitudes of more able 9 and 13-year-olds are also salient for 10 and 11-year-old children across the ability range. A major concern of this study is that if computer-mediated assessment is to be a valid and reliable alternative to paper-based assessment then it is important that developers of computer mediated assessments are aware of the effect of mode on children's behaviour.

Through the analysis of children's methods and behaviour this study hopes to develop an understanding of how children think when working in the two different modes. In conjunction with the findings of Johnson & Green (2004), comparisons of performance and behaviour in the

two modes can lead to inferences about how working in the different modes may have affected the children's mental processes.

Methodology:

Two tests were constructed (Appendix 1). Each test contained 10 mathematics questions spanning National Curriculum levels 3, 4 and 5. The tests were administered to 104 Year 6 (10 and 11-year old) children in both paper-based and computer-based formats. The children were selected from four primary schools - one large urban school, one small urban school, one large suburban school and one small suburban school. All of the children in participating classes were invited to take part in the study and those gaining parental consent were included.

The children were put into four experimental groups, being randomly assigned to these groups from a sampling frame constructed from lists provided by each of the schools. This was done so that each school had an even number of children and an even gender split within each of the experimental groups, as far as possible. Tests were allocated so that approximately half did Test A first and half did a paper-based test first.

	1st test	2nd test	n
Experimental group 1	Test A paper	Test B computer	27
Experimental group 2	Test B paper	Test A computer	26
Experimental group 3	Test A computer	Test B paper	26
Experimental group 4	Test B computer	Test A paper	25

The questions for each of the tests were matched for difficulty according to their National Curriculum criteria and level. Each test contained two questions from level 3, six from level 4 and two from level 5. The questions were selected according to a number of criteria. Questions which gave children the opportunity to make their working processes explicit were chosen so that observations could be made about how they approached the problem. Choosing questions which demonstrated a variety of characteristics was also a consideration, e.g. the response types, the use of tools, the number of 'steps' involved, the level of contextualisation, the type of operation involved. Children were provided with a blank sheet of paper on which to show their working. This was done to encourage children to show their working and to facilitate the collection of working method data for analysis.

Issues relating to school based access to the internet site hosting the questions and children's ease of navigation through the tests were investigated in a pilot study prior to the main study. This also allowed researchers to test out and refine observation and interview techniques. As a result of the pilot study two questions which had been initially chosen for the tests were changed. One was considered to be too demanding for children who would be new to Year 6, whilst the other relied on mathematical symbol conventions which the children in the pilot study felt uncomfortable with.

An earlier study (Johnson & Green, 2004) had explored whether the children had performed differently in the tests according to mode. In that study facility values were analysed to explore the impact of the mode on performance. Errors were coded and this allowed further investigation of the differences between questions in the different modes.

The aim of the present study was to supplement this performance data with extra information gathered whilst observing the behaviour of a sample of children as they completed the tasks. The rationale for doing this was to gain an insight into the effect of motivational factors and capture children's affective responses to working on computer and paper

A sample of 8 children was drawn from 4 schools, representing a mix of ability and gender. Observation of the children attempting questions in both modes was facilitated by the use of a structured, pre-designed observation schedule. Cohen, Mannion & Morrison (2000) suggest that this is a useful tool if the observation's concern is 'to chart incidence, presence and frequency of [interactional] elements' (2000, p.306). The final study schedule (Appendix 2) included a variety of low inference measures, specific identifiable behaviours based on feedback from a pilot study (e.g. 'reading aloud' or 'referring backwards and forwards'), with some high inference measures relating to more global reasons for behaviour (e.g. indications of 'boredom', 'frustration' or 'relaxation').

In order to address concerns about subjectivity and interpretation, interviews with the eight sample children were used to help confirm or refute an observer's interpretation of motives for particular types of behaviour. The structure of interviews was designed to enable children to verbalise their working methods. Children were shown the matched tasks and asked to describe if there were any differences in the way that they worked out each of the problems.

This process of Stimulated Recall (SR) was facilitated by the use of a replay option in the computer software that allowed children to see their response and any revisions that they made during the answering process. Children were also shown any working jottings that they

may have made whilst answering questions. Whilst SR has been used extensively as a method to trace cognitive activity in educational research, a number of concerns have been raised about its validity (Yinger, 1986; Tjeerdsma, 1997). Taking into consideration these criticisms, Lyle has argued that 'there is little doubt that education can benefit from SR research' (2003, p.874) so long as great care is taken regarding SR procedures. Lyle suggests that for the method to be reliable and valid it is necessary to ensure that there is a minimal time delay between event and recall. Furthermore, a researcher's questions must not alter the cognitive process being recalled by suggesting ideas to be considered by the subject.

Taking these points into consideration, interviews were held with children following the completion of their second task, therefore minimising the time delay between event and recall. Furthermore, during the subsequent interview the researcher did not prompt children unless an event that occurred during task observation needed clarification.

During the interviews it was also possible to gather data about children's familiarity with computers and levels of use at home since it was felt that this could be useful for later analysis. Finally, the use of a number of observers throughout the study was intended to help negate any dominant assumptions that may have underpinned interpretations made by a single observer.

Another important strand of this investigation involved analysing children's working methods, providing another insight into children's thought processes as they completed their tasks. Working methods were coded where a child used a different strategy in matched questions in different modes.

Findings

Strategies:

Children's working methods were gathered and analysed. We were able to compare strategies for 83 children who showed working for both modes for at least one question. 39 of these children changed their strategy according to mode. This means that they chose a different working method when attempting questions that were based on common criteria but where one was attempted on paper and one on computer. Whether the student got both, one, or neither of the matched questions correct was not considered to be important, since the focus of the study was to capture evidence of process rather than performance. Performance differences between modes was covered in the first phase of the study (Johnson & Green, 2004).

Girls were over-represented in the group of children who applied different methods to matched questions, suggesting that they were more likely to alter their strategy according to question instance or mode. This finding was statistically significant (Appendix 3).

Furthermore, girls were also over-represented ($^{14}/_{19}$) in the group of children who changed their method in two or more different matched questions. This implied that girls were not only more likely to change their method when attempting matched questions, but that they were also more likely to do this more than once within the given 10 questions.

Although the number of children who changed their method according to matched question was relatively small it was still possible to discern patterns within some of the questions. For questions that asked children to add two numbers (352+39; 472+18) it was more common for children to adopt a standard written method when working on computer. Of the 8 children who changed their method, 5 chose to use a standard written addition method for computer versions of the questions whilst the same number chose to use partitioning strategies when attempting matched questions on paper.

This tendency to use partitioning on paper rather than on computer was mirrored in data from questions that asked children to subtract one number from another (554-538; 546-39). For these questions 5 of the 11 children chose to use partitioning strategies when attempting the questions on paper whilst only 2 used this strategy on computer.

For one of the most difficult questions there appeared to be an interesting mode-related influence on children's strategies. Question 5 was set out as a standard written addition problem where children were expected to fill in missing digits (*xx*+89=*x*43; *xx*+*58*=x11). All 7 of the children who chose a different strategy according to mode adopted an addition based approach to solving the problem on computer whilst 3 of these children chose to use an inverse subtraction method on paper.

For the question that asked children to calculate the perimeter of an irregular shape, mode appeared to have an influence on the strategies chosen. Of the 9 children who changed their method according to mode, 7 chose a cumulative approach on paper. This meant that they tended to add the measures for each individual side of the shape as they counted around the shape (e.g. 20+4+8+12+8+4...). On the other hand, 6 of the children chose a combination strategy when working on screen. This meant that they tended to group the numbers relating

to matching sides together before combining all of the numbers into a total (e.g. 20+20+20=60, 8+8=16, 4+4=8...).

Perceptions:

Overall perceptions

A sample of 8 children was drawn from four schools, representing a mix of gender and ability. The level of home computer use within the sample varied. One child had no home computer access whilst the others generally spent between 30 minutes and an hour per day using a home computer.

When asked about their initial feelings about answering questions on computer most of the children felt that it was a favourable experience. This mirrors findings by Choi & Tinkler (2002). Many preferences for particular questions were made on the basis of the specific numbers involved in the questions and were therefore not mode-related. Where mode-related preferences were found they were largely generic, being relevant to a number of different questions.

Of the seven children who gave mode-related reasons for their answer two children liked using computer-based tools and having less writing to do. One child felt that he paid more attention to computer-based questions, and another child thought computer-based questions were less difficult than paper-based questions. Only one child felt that computer-based questions contained an element of difficulty which paper-based questions did not. He suggested that computer-based questions restricted his working because he couldn't write his working down as easily on screen. Despite this he still had a positive attitude to answering questions on computer. Two of the eight children felt that the experiences of answering computer-based questions and paper-based questions were similar.

In favour of computer

5 of the children felt that computer-based questions were easier than paper-based questions. The greatest generic reason for preferring computer-based questions was the use of keys and the need for less writing. Most of the children felt that 'using keys', 'using tools' and 'doing less writing' made questions easier.

Other reasons related to question layout being clearer on screen and a sense that computer-based questions 'lead you through' the test. A number of comments also related to the idea

that computer-based questions were more enjoyable than paper-based questions, with one child suggesting that 'boring content' could be more fun when presented on screen. The same child also felt that paper-based tests implied time limits, unlike computer-based tests. Finally, one child also felt that having to show working out on the question page led to a cluttered and confusing appearance.

An interesting finding was that questions that had no contextualisation were more likely to be preferred on computer.

In favour of paper

Only two of the eight children felt that computer-based questions were more difficult than paper-based questions. The greatest generic reason for preferring paper-based questions related to not having to transfer attention from page to screen when working out problems. A number of children said they liked their working to be near to the question so that they did not have to look away from the problem. These children suggested that switching attention from page to screen to refer to notes contributed to a sense of difficulty, whereas paper-based questions provided a natural space to show working out. The affordance of having space on the page was mentioned as being important for one child who liked to write numbers over the text in one of the contextualised paper questions to support his working.

The use of the on-screen protractor was also mentioned as a source of difficulty, specifically the manipulation of the protractor around the screen. Angle measuring questions were generally preferred on paper, especially those involving larger angles that required rotation of the protractor.

Finally, perimeter calculation questions were generally favoured on paper.

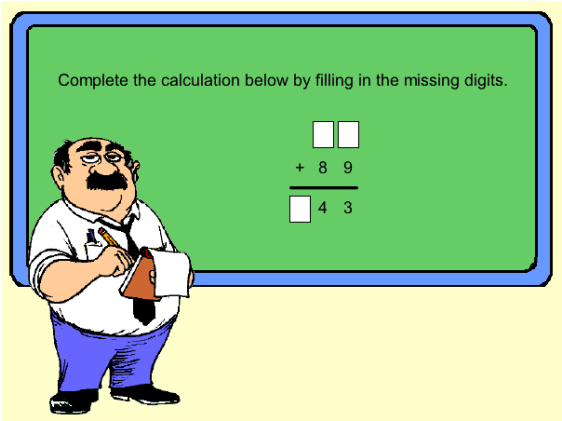
Specific question-related perceptions

Although the small sample size makes it important not to generalise too confidently, some mode-related findings emerged when children expressed preferences for specific questions. In 8 of the 20 questions it was clear that the mode affected preferences and in these cases the preferences were not related to the content of the questions (Appendix 4).

In favour of computer

Specific mode related preferences for particular computer-based questions were found in two questions.

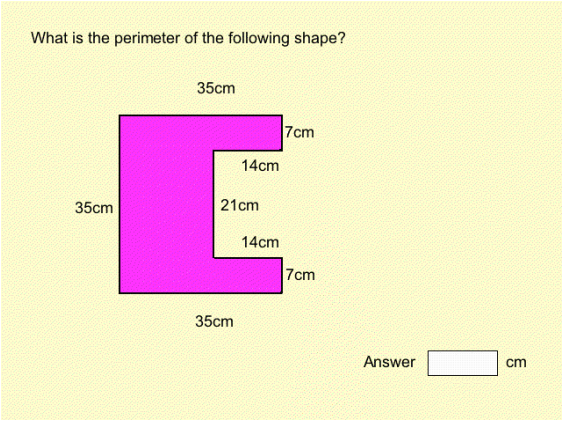
One child felt that the screen version supported the context of the question because the colours on screen matched the hair colouring referred to in the question text. This was not as clear on the black and white paper version of the question.



One child thought that the green background in the screen version was bolder than in the paper version. He felt that this helped to emphasise the space that he needed to fill. He felt that this was important because he failed to fill one of the boxes on the paper version of the question and subsequently submitted an incorrect answer.

In favour of paper

Specific mode related preferences were only found in one particular paper-based question.



One child said that he liked having the shape provided on paper so that he could count around that perimeter without missing any sides out of his calculation. This mirrors the finding in the *strategies* analysis where most children chose a cumulative strategy when attempting to solve the problem on paper.

The angle measuring questions were problematic. Mode affected student perceptions in a variety of ways in these questions. Half of the sample ($^4/_8$) preferred the paper-based versions of the questions because they felt it was easier to rotate the angle by moving the page and

there was no need to crane their necks in the process. Others felt that it was easier to position the manual protractor compared with the on-screen protractor. Finally, one child felt that the manual protractor was visually clearer than the on-screen version.

Of the other half of the sample who preferred the computer-based versions of the angle questions most comments related to the 'fixed' nature of the on-screen protractor. One child felt that the computer protractor stayed more still and 'wobbled less' than a manual protractor, whilst another felt that it was less difficult to position. Another child liked the way that the on-screen protractor could not be placed on the angle 'upside down' since its orientation was correct by default. A final comment suggested that the tool introduced an element of 'fun' into the question, leading them to pay more attention to the problem.

Behaviour:

In all cases children completed their paper test more quickly than their computer test. There was only one exception, where one child took an equal length of time for both tests.

'Off task' behaviours were slightly more common on computer and differed in nature from behaviours observed during paper tests. 3 of the 8 children were prone to distraction whilst questions loaded onto their computer. On the other hand, distractions during paper tests tended to be caused by distractions elsewhere in the room, such as sudden noise or movement. This type of distraction was also noted during computer tests but to a lesser extent.

A number of children exhibited mode-related behaviours when completing the angle measuring questions. Half of the children showed signs of craning their necks whilst working on computer but not on paper. 3 of the 8 also appeared to struggle to read the on-screen protractor but not the manual protractor. 5 of the children adopted a strategy of rotating the paper rather than the protractor when attempting one or both of the angle questions on paper.

There were some interesting mode-related differences in the way that some children moved through their tests. The paper test involved a greater degree of flexibility than the computer test because children could move forwards or backwards whenever they felt it necessary. This was not possible on the computer where a submitted answer led children automatically to their next question. Three of the children checked through their answers at the end of their paper test, and made some amendments. Two of the children looked on to forthcoming questions when partway through their paper test. Of these, one child 'previewed' the next question to decide whether to move on and skip the question he had just reached. Finally, one child

showed relief and some enjoyment when turning over the last page of their test booklet to reveal the end of the test.

Discussion

Strategies

The findings of this study suggest that mode affected strategy choice for around 37% of the children overall. Furthermore, mode appeared to affect girls more than boys, with girls tending to use more flexibility when applying strategies between modes.

Findings also appeared to suggest that children tended to have a more flexible approach to problem solving on paper. When working on computer children were more influenced by the way that the question was physically presented. This effect was evident in Question 1 (Appendix 1) where children tended to approach the problem on computer by using a standard written addition strategy. On the other hand, when working on paper children were less likely to use this formal strategy, instead tending to use informal partitioning strategies. This pattern was mirrored in Question 5 (Appendix 1) which was physically laid out in the form of a standard column addition problem. The children who altered their strategies between modes chose to solve this problem on computer using an addition process, reflecting the manner of its presentation. When attempting the matched problem on paper the most common strategy was inverse subtraction – possibly a more effective approach to dealing with the problem.

Although this study involved relatively small numbers, it appears that there was a group of children who had a tendency to interpret and act on screen-based problems differently than paper-based problems. Furthermore, it appears that some of the children in this group were more likely to apply more flexible strategies to paper-based problems. One suggested reason for this may be that for some of the children there might be a tendency to view objects presented on screen as being more 'fixed' than those presented on paper. If questions presented on screen are taken 'at face value', i.e. problems presented in an 'addition' format imply addition strategies, it is possible that alternate and possibly more effective strategies may be overlooked.

It could be argued that a difference in perception between data presented on screen and on paper may relate to common classroom experience. It is more likely that children will experience mathematical processes involving thinking around problems, manipulating numbers and offering alternative solutions on paper rather than on screen. Furthermore, it could be

inferred that this practice will be more common in the primary school years where access to computers is more limited (D*f*ES, 2003) than in secondary schools. This possible connection between common classroom practice, perceptions of screen-based problems and strategy use may help to explain the findings of Choi & Tinkler (2002) and Coon et al (2002) who suggested that primary-aged children found computer-based questions more difficult than paper-based questions.

The impact of mode on strategy choice was very apparent for the perimeter calculation questions. The evidence suggests that children were more likely to use a cumulative approach when working on paper and it appears that the affordances of the paper medium promoted this strategy. The data shows that children used a more tactile approach to solving the problem on paper, 'ticking off' or 'dotting' each number around the shape as they accommodated it into their calculation. This approach did not translate into the computer medium where children tended to mentally combine numbers together before calculating the total. The inclusion of this extra 'combination' step in the process may have led to a number of children failing to reach an answer on the computer-based perimeter questions. It is interesting that this error type was not found in the matched paper-based perimeter questions.

The effect of the affordances of mode on strategy choice appears to support the work of Greenwood et al (2000) who found that computer-based spatial awareness questions were more difficult than paper-based questions for secondary-aged students. The findings of this study suggest one reason why such an effect may also extend to children within the primary school sector.

Perceptions

The interview data show that the majority of children felt computer-based questions were easier than paper-based questions. This is an interesting finding when compared with the performance data from phase 1 of the study (Johnson & Green, 2004) since the empirical evidence suggested that computer-based questions were often more difficult than paper-based questions. The notion of ease may be a consequence of both the technical affordances of the computer medium and other perceptual issues connected to children's experiences of computers in the wider environment.

There were a number of physical and technical features of computers that affected children's preferences. For most of the children in this study the concept of 'task ease' was related to 'doing less writing'. An obvious feature of the computer medium is the facility to use keying or

tools, thereby avoiding manual written activity. As a consequence it appears that computer technology has an in-built advantage over the paper medium since it avoids a crucial area that appears to contribute to children's perceptions of difficulty. This finding supports those of Richardson et al (2002) who reported a similar reaction from higher ability 9 and 11-year-olds. Other layout features such as the use of colour and the combination of coloured graphics with supporting text were also felt to make questions easier on computer. Again, these findings are in agreement with those of Richardson et al.

It is possible that reasons for preferences are also influenced by broader attitudes about computers which affect children's perceptions. The work of Levin & Gordon (1989) and the findings of phase 1 of this study (Johnson & Green, 2004) both support the idea that children may link the activity of answering questions on screen with prior experiences in the medium and other activities commonly associated with computers, namely games. It could be argued that the influence of a 'computer game schema' might influence children's perceptions about the true demand of computer-based questions. The logic of this argument implies that a positive disposition to working in the computer medium may lead to a perception that questions presented on computer may be less demanding than those presented on paper. This could be an important finding since it suggests that children may have a more positive attitude and in turn greater motivation to complete computer-based questions than paper-based questions.

It is also possible to suggest that overall positive attitudes to computer-based questions mask some specific question features where there was not a positive reaction. Although children generally preferred answering questions on computer there was a group of questions where this trend was reversed. Children preferred shape, space and measure questions on paper, supporting the findings of Greenwood et al (2000). It is interesting to note that the specific demand to integrate spatial awareness, tools and calculations had a greater bearing on children's perceptions than did the generally positive disposition towards the computer medium.

Behaviour

It appeared that mode affected children's on- and off-task behaviour. Whilst answering questions on computer children showed a greater commitment to the task, being less prone to extraneous influences and distractions. This suggests that the children were more motivated and engaged with the problems presented on screen, an engagement not found to the same extent on paper.

The greater involvement with computer-based questions needs to be balanced with differences in off-task behaviour between the modes. One consequence of working in the computer medium was the tendency for children to lose focus whilst questions loaded. The level of distraction appeared to be linked directly to download speeds, with slow functioning systems leading to more instances of off-task behaviours. This has implications for teachers and classroom management. It might be suggested that where large groups of children are simultaneously working on-line there could be a direct relationship between a school's technical capacity and classroom ambience.

There were differences between the way children behaved between modes. At times these differences resulted from the limitations of computer testing, therefore we were not comparing like with like. For this study questions were delivered on screen without the opportunity for children to go back and review previously submitted answers. This design feature was planned into the computer test to observe the effect of the difference between modes, since children were only able to review past questions (and preview future questions) in the paper medium. This allowed the study to focus on the way that the children navigated through their test questions to see if evidence from observations would support the findings of Greenwood et al (2000) who suggested that secondary-aged students tended not to review their work. Observation evidence found that children did tend to review and amend their answers when they had the opportunity. Furthermore, some children navigated through their paper tests by previewing forthcoming questions, apparently 'weighing up' whether to attempt some questions before others. It was also possible to observe children reviewing past strategies to inform their approach for new questions. Such observations indicate that the children were seeing the test questions in relation to each other, and that mental processes used were not considered redundant after the closure of an individual question. It appears that children were given a degree of independence and control on paper which they didn't possess on computer and which allowed them access to strategies that could facilitate their performance.

Another difference between the way children behaved according to mode was found within the angle measuring questions. Whilst the computer software only allowed the protractor to be manipulated, observations of paper-based behaviour showed a number of children manipulating the paper rather than the protractor. This is another example where the affordances of the technology limited the opportunity for some children to behave on screen as they would on paper. It could be argued that these technical limitations (i.e. that you measure angle by manipulating a protractor) may discriminate against some children who do not conform to those behaviours.

The findings of this study suggest that understanding perceptions is a complex issue. It does appear that certain types of questions in certain domains have different impacts according to mode. It is important that those involved in the developing shift towards online testing are aware of any effects inherent in this shift. It is also important that the limitations of technology are not allowed to negatively influence the way that children behave when answering questions within a test. If limitations are built into test and question designs (i.e. the inability to review past questions or the inability to rotate the angle) then it is important that the effects of such limitations are known so that they can be taken into account.

Bibliography:

Bearne, E. & Kress, G. (2001) Editorial. *Reading, Literacy and Language* 35 (3) 89-93

Bennett, R.E. (2003) Online Assessment and the Comparability of Score Meaning. *Paper presented to International Association for Educational Assessment Annual conference, Manchester, October 2003*

Bridgeman, B., Lennon, M.L. & Jackenthal, A. (2002) Effects of screen size, screen resolution, and display rate on computer-based test performance. *Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans, April 2002*

Choi, S. W. & Tinkler, T. (2002) Evaluating comparability of paper-and-pencil and computer-based assessment in the K-12 setting. *Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans, April 2002*

Cohen, L., Manion, L. & Morrison, K. (2000) *Research methods in education*. London: RoutledgeFalmer.

Coon, C., McLeod, L. & Thissen, D. (2002) *NCCATS update: Comparability results of paper and computer forms of the North Carolina End-of-Grade Tests (RTI Project No. 08486.001)*. Raleigh, NC: North Carolina Department of Public Instruction.

D*f*ES (2003) *Information and Communications Technology in Schools in England: 2003*, available from http://www.dfes.gov.uk/rsgateway/DB/SFR/s000405/index.shtml

Greenwood, L., Cole, U.M., McBride, F.V., Morrison, H., Cowan, P. & Lee, M. (2000) Can the same results be obtained using computer-mediated tests as for paper-based tests for National Curriculum assessment? *Proceedings of the International Conference on Mathematics/Science, Education and Technology*. Vol. 2000 (1) 179-84

Johnson, M. & Green, S. (2004) On-line assessment: the impact of mode on student performance. *Paper presented at British Educational Research Association Annual Conference, Manchester, September 2004*

Levin, T. & Gordon, C. (1989) Effect of gender and computer experience on attitudes toward computers. *Journal of Computing Research* 5 (1) 69-88

Lyle, J. (2003) Stimulated Recall: a report on its use in naturalistic research. *British Educational Research Journal* 29 (6) 861-78

Raikes, N. & Harding, R. (2003) The horseless carriage stage: replacing conventional measures. *Assessment in Education* 10 (3) 267-77

Richardson, M., Baird, J., Ridgway, J., Ripley, M., Shorrocks-Taylor, D. & Swan, M. (2002) Challenging Minds? Students' perceptions of computer-based World Class Tests of problem-solving. *Computers and Human Behaviour* 18 (6) 633-49

Tjeerdsma, B. L. (1997) A comparison of teacher and student perspectives of tasks and feedback. *Journal of teaching in physical education* 16 388-400

Yinger, R. J. (1986) Examining thought in action: a theoretical and methodological critique of research on interactive teaching. *Teaching and teacher education* 17 725-40

Appendices

Appendix 1

	Test A	Test B
1	There are 472 boys and 18 girls at the cinema. How many children are there altogether?	There are 352 boys and 39 girls at the cinema. How many children are there altogether?
2	Ann scored 554 points in a computer game and Alan scored 538 points. What is the difference in their scores?	Mary scored 546 points in a computer game and Fiona scored 39 points. What is the difference in their scores?
3	At an antique doll fair there are 25 dolls with black hair, 21 dolls with brown hair, and the remaining dolls have fair hair. If there are 90 dolls on display, how many have fair hair?	At an antique doll fair there are 32 dolls with black hair, 18 dolls with brown hair, and the remaining dolls have fair hair. If there are 70 dolls on display, how many have fair hair?
4	Vera went shopping with £70 to spend but only spent £49. She put the rest of the money into her savings account which already had £350 in it. What was the final amount of money in the savings account?	Gavin went shopping with £84 to spend but only spent £43. He put the rest of the money into his savings account which already had £399 in it. What was the final amount of money in the savings account?
5	∀∀ + 5 8 ∀ 1 1	∀∀ + 8 9 ∀ 4 3
6	∀∀ - 2 6 2 9	∀∀ - 4 5 3 6
7	Bob plants 15 rows of turnips in his vegetable garden. There are 25 turnips in each row. How many turnips does he plant?	David plants 15 rows of carrots in his vegetable garden. There are 13 carrots in each row. How many carrots does he plant?
8	What is the perimeter of the following shape? (20cm+20cm+20cm+4cm+8cm+12cm+8cm+4cm)	What is the perimeter of the following shape? (35cm+35cm+35cm+7cm+14cm+21cm+14cm+7cm)
9	Use the protractor to measure the angle below. (36^0)	Use the protractor to measure the angle below. (28^0)
10	Use the protractor to measure the angle below. (65^0)	Use the protractor to measure the angle below. (50^0)

Appendix 2

Observation schedule

Pupil name: _____ ***Mode:*** paper ***Time started:*** _____

A: Spaces within the question

On task behaviour

reading aloud/mouthing question	
thinking aloud process	
shifting sitting position	
hovering/flicking over boxes with pencil	

Off task behaviour

looking away from paper – engaged elsewhere	
shifting sitting position	
hovering/flicking over boxes with pencil	

B: Movement through the test

move on without answering	
referring backwards and forwards	
checking answers at end	
going back to unfinished questions	

C: Angle questions 9 & 10

rotating paper	
craning neck	
struggling to read the numbers	

Any other observations?

……..
……..
……..

Record time finished: _____

Signs of affective response			*Notes about behaviour*
relaxed	1 2 3 4	tense	
satisfied	1 2 3 4	frustrated	
interested	1 2 3 4	bored	

Appendix 3

method difference * pupil gender Crosstabulation

Count

		pupil gender		
		boy	girl	Total
method difference	method difference	13	26	39
	no method difference	46	19	65
Total		59	45	104

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	13.916[b]	1	.000		
Continuity Correction[a]	12.433	1	.000		
Likelihood Ratio	14.090	1	.000		
Fisher's Exact Test				.000	.000
Linear-by-Linear Association	13.782	1	.000		
N of Valid Cases	104				

a. Computed only for a 2x2 table

b. 0 cells (.0%) have expected count less than 5. The minimum expected count is 16.88.

Appendix 4

Preferences for question instances by mode

Question instance	N° of children preferring computer version	N° of children preferring paper version
472+18 (contextualised)	**1/3**	**3/5***
352+39 (contextualised)	2/5	1/3
554-538 (contextualised)	1/3	2/5
546-39 (contextualised)	**2/5**	**2/3***
90-(25+21) (contextualised)	0/3	2/5
70-(32+18) (contextualised)	3/5	3/3
350+(70-49) (contextualised)	1/3	2/5
399+(84-43) (contextualised)	**2/5**	**2/3***
xx+58=*x*11 (not contextualised)	1/3	1/5
***xx*+89=*x*43 (not contextualised)**	**4/5***	**1/3**
***xx*-26=29 (not contextualised)**	**2/3***	**2/5**
xx-45=36 (not contextualised)	2/5	1/3
15x25 (contextualised)	1/3	2/5
15x13 (contextualised)	3/5	2/3
Measure perimeter (total 96cm)	**1/3**	**3/5***
Measure perimeter (total 168cm)	1/5	1/3
Measure 36^0 angle (no protractor rotation necessary)	1/3	2/5
Measure 28^0 angle (no protractor rotation necessary)	3/5	2/3
Measure 65^0 angle (some protractor rotation necessary)	**1/3**	**3/5***
Measure 50^0 angle (some protractor rotation necessary)	**2/5**	**2/3***

Bold font shows instances where sample preferences were split more than half/less than half with *showing the majority preference