

The final publication is available at [www.springerlink.com](http://www.springerlink.com)

DOI: <http://dx.doi.org/10.1007/s10798-011-9191-3>

## **Assessing Scientific and Technological Enquiry Skills at Age 11 using the e-scape System**

**Dan Davies, Chris Collier and Alan Howe**

Bath Spa University

### **ABSTRACT**

This article reports on the outcomes from the *e-scape Primary Scientific and Technological Understanding Assessment Project* (2009-10), which aimed to support primary teachers in developing valid portfolio-based tasks to assess pupils' scientific and technological enquiry skills at age 11. This was part of the wider 'e-scape' project (2003-present), which has developed an innovative controlled alternative to design & technology and science public assessment at age 16. Teachers from eight primary schools were trained in the use of an online task-authoring tool to develop and trial assessment activities based on current classroom work. To compile their e-portfolios of assessment evidence, pupils used netbook devices, which afford multi-modal responses (text, drawing, photo, audio, video, spreadsheet) whilst leaving space on pupils' tables for practical investigations. Once the pupil e-portfolios had been uploaded to the secure e-scape website, teachers assessed them using a 'comparative judgement' approach to produce a rank order with a high reliability coefficient. Participant teachers recognised the strength of the e-scape approach in terms of facilitating and managing pupils' responses to assessment tasks in the classroom, which they successfully adapted to suit primary pedagogy. In particular, the benefits of scaffolding complex assessment tasks through the step-wise e-scape process in the authoring tool represented for some of the teachers a pedagogically significant development in terms of their planning.

### **KEYWORDS**

Assessment, primary, e-portfolio, science, technology

## INTRODUCTION

This article reports on an e-scape project which sought to solve one of the key issues in assessment for primary teachers – namely, how to assess children’s enquiry skills in ‘real time’ (whilst they are actually undertaking an enquiry activity in the classroom). This issue has become particularly important in primary science education in England since the discontinuation of the Standard Attainment Tests (SATs) which all 11-year-old pupils took in this subject until 2009. Since this date, pupil performance data in science reported to the UK government by each primary school has relied entirely on teacher assessment undertaken in the classroom. Although this arguably provides a more valid picture of children’s scientific enquiry skills than a paper and pencil test, observations of science investigations in progress face the problem that the teacher cannot be everywhere at once. Whilst the teacher is listening in to one group, elsewhere in the class there may be some significant learning going on that has been missed. Such learning is not always captured in the ‘write up’ of the investigation either; some children who can think well scientifically have difficulty in expressing their ideas in writing. For many children – and teachers – the reliance on written evidence for assessment is the least attractive aspect of science, particularly at the upper primary level.

Another recently planned change affecting primary teachers in England was the *Independent Review of the Primary Curriculum* (Rose 2009). This proposed to combine science with design & technology (D&T) to create a new ‘learning area’ of scientific and technological understanding, further adding to the potential complexity of assessing pupils’ enquiry skills in this composite discipline. For this reason, the Centre for Research in Early Scientific Learning (CRESL) at Bath Spa University decided to work with a group of specialist teachers in science, D&T and ICT in eight primary schools to develop e-scape tasks to assess 11-year-old pupils’ scientific and technological enquiry skills. The approach we took built upon the findings of part of the *e-scape Phase Three* project (Kimbell *et al* 2009), which sought to develop e-portfolio assessment of science at age 15 (Davies 2009). This found that three-hour assessment tasks (typically one morning) were sufficient to complete a science investigation, although the proposed incorporation of some D&T elements in each task could potentially expand this time-frame. We also drew upon the findings from *e-scape phase 2* (Kimbell *et al* 2007) in the development of D&T tasks and the subsequent expansion of e-scape into primary schools during 2008-9; we visited one of these schools before starting our project to observe a task on

'bungee jumping' combining elements of scientific and designerly enquiry. The model of assessment we proposed to use for the *e-scape Primary Scientific and Technological Understanding Assessment Project* (2009-10) was an improvement on standard approaches to primary classroom assessment of enquiry skills in the following respects (Table 1):

**Table 1: Comparison of standard primary classroom techniques with e-scape portfolio enquiry**

<b>Standard primary scientific enquiry</b>	<b>e-scape portfolio enquiry</b>
Investigation prompts come from teacher/whiteboard/planning board/worksheet	Investigation prompts on screen (children follow prompt instructions or questions, record as directed then move onto next box)
Record of enquiry written up after investigation	Recording of ideas and outcomes takes place within enquiry time
Usually part of investigation selected to be written up	Recording can take place at every stage of investigation (each box can contain prompt questions and then a choice of methods to record answers)
Hand written recording	Choice of text, voice recording, photos, drawings, video
Use of ICT (to graph etc) done separately during class slot in computer room	Results recorded directly onto spreadsheet, which can be graphed immediately
Investigate as a group, prepare group presentation or record on own	Investigate as a group, record as a group (especially voice recordings and video) or record individually (depends on number of fizzbooks/computers with microphones and cameras)
Different recording methods will be in different places, e.g. write up in books, graph on PC, photo on teacher's camera, group presentation on video camera...	All outcomes in one place – in e-portfolio

## **ASSESSMENT OF ENQUIRY SKILLS AT PRIMARY LEVEL**

There is a clear consensus within the literature that assessment is an activity integral to learning and teaching science in the primary school and that both children's conceptual development and their ability to carry out scientific enquiries should be assessed (Ward *et al* 2005; Harlen 2007; Howe *et al* 2009). While these and many other science educators advocate the formative and summative assessment of science enquiry skills during practical classroom activity, such an approach was not used in national statutory end-of-primary tests between 1996 and 2009. SATs assessed knowledge and understanding of scientific inquiry in a summative, atomistic way, through presenting brief outlines on paper of invented scenarios intended to draw the child into a classroom context before asking them a series of questions about a fictional science inquiry. This approach has been criticized as unreliable, with up to 13% of 11 year olds being assigned to the wrong level in science (He, Hayes

and Wiliam, 2011). Wyse *et al.* (2008) identified the negative effects of the implementation of national curriculum assessment requirements on practical scientific investigation, which many primary teachers came to see as a time-consuming classroom activity which could not be accommodated when children were revising for SATs. An intense focus on testing and test results in the core subjects of English, Maths and Science narrowed the curriculum and drove teaching 'in exactly the opposite direction to that which research indicates will improve learning and attainment' (*op. cit.* p. 18). Furthermore, primary science SATs were perceived to be 'one of the biggest obstacles to delivering high-quality, engaging education at this level' (Harlen 2008, p. 3) and research by Collins *et al.* (2008 cited in Tymms *et al.* 2010) suggests that the recent abolition of testing in science at Key Stage 2 in Wales is having a beneficial effect on the development of 11-year-old children's knowledge and understanding of science.

Since end of Key Stage 2 SATs were abolished in England, summative assessment in science has become the teachers' responsibility. Harlen (2008, p. 16) recommends that there should be 'no high-stakes use of summative assessment of pupils' progress' at the primary level as this distorts teaching and learning, and that national tests should be replaced by moderated teachers' assessment, so that progress in the full range of skills and concepts can be recorded and reported. She proposes that

....there are several ways of raising the reliability of teachers' assessment. The examples of practice in various countries show that the most commonly used are group moderation and the use of special tests or tasks that have been tried out and calibrated for teachers to use to check their judgements.

(Harlen 2008, p. 2)

Harlen and Deakin-Crick (2002, p.4) have found that

'[w]hen passing tests is high stakes, teachers adopt a teaching style which emphasises transmission teaching of knowledge, thereby favouring those students who prefer to learn in this way and disadvantaging and lowering the self-esteem of those who prefer more active and creative learning experiences'.

So, rather than replicate SATs-style summative assessment, the development of new models of moderated teacher assessment to assess pupils' progress in the full range of enquiry skills is required.

Tensions clearly exist between authenticity and manageability in the design of tasks to assess children's enquiry skills. Harlen (1999, p.137) suggests therefore that 'special assessment tasks' should be made available to allow pupils to show the skills that they have. The combination of a summary of ongoing assessment and some well-designed practical tasks is judged by Harlen to be 'the best compromise' for the assessment of practical skills. She goes on to describe how information can be gathered in the form of observations, questioning, specific task setting and asking pupils to 'communicate their thinking through drawings, artefacts, actions, role play and concept mapping, as well as writing' (p. 133). Harlen and Deakin Crick (2002) recommend strategies to reduce the observed negative impact of summative assessment on motivation for learning, including:

- avoiding drill and practice tests;
- de-emphasising tests by using a range of forms of classroom assessment and recognising the limitations of tests;
- avoiding children being faced with tests in which they are unlikely to succeed (Harlen and Deakin Crick (2002, p. 70).

It is sometimes difficult to draw clear distinctions between summative and formative assessment strategies. Hodgson and Pyle (2010) propose that assessment for learning (AfL) strategies can involve using summative tests for formative purposes. They go on to note that assessment does not have to include a measurement. Rather, it can take the form of purposeful enquiry that can potentially enhance pupils' learning. Desirable pupil outcomes attributable to AfL include 'authorship of their own learning' and pupils 'constructing knowledge collaboratively' (Hodgson and Pyle 2010, p.3). Both theory and practice suggest that the main elements of effective enquiry-based classroom practice which can be incorporated into AfL occur when...

- teachers provide the means for children to collect evidence, which may be through experiment and practical inquiry or from secondary sources;
- children have the opportunity to express their ideas, to listen to the ideas of others and to build on their existing ideas when faced with new experiences. This means they have shared experiences to discuss, time to do this and, where appropriate, real objects to handle and explore;
- teachers pose questions that require children to hypothesise, predict and suggest answers;
- teachers engage children in thinking about and discussing how to test their predictions and see if their ideas 'work';
- children are clear about what they are finding out and what they are learning by doing so;
- children consider the evidence they collect in relation to initial ideas and predictions;
- children reflect and report on how and on what they have learned. (Harlen 2008, pp. 13-14)

It is clearly difficult to capture in meaningful ways the details of children's thinking and actions as they work through a carefully designed and relatively complex assessment task. The assessment of children's discussions, hypotheses and reflections within the context of group work in a primary classroom presents a particular challenge. It is becoming apparent that the affordances of Information and Communication Technology (ICT) can help teachers meet this challenge. Murphy (2006, p.19) believes ICT can 'greatly enhance' opportunities for children to communicate their scientific thinking and thereby enable them to better make meaning within a social constructivist teaching methodology. Gillespie (2006) identifies virtual learning environments as having potential to support investigative science by enabling children to interact, communicate and collaborate. E-portfolios have the potential to provide a real-time, authentic account of a learner's journey through an active learning task/design challenge (McLaren 2010). They can include a series of software tools to enable learners to present their emerging ideas using a variety of media. In addition, they can offer prompts for '... peer and self evaluations, reflection, review and target setting' (McLaren 2010, p. 3). The resulting e-portfolios have the potential to facilitate moderated teacher assessment as they can be viewed to explore the detail of learners' work. The literature suggests that e-portfolios can have multiple purposes:

'as assessment tools to document the attainment of standards (a positivist model--the assessment portfolio); as digital stories of deep learning (a constructivist model--the learning or process portfolio); and as digital resumes to highlight competence (a showcase model-- the best works/marketing/employment portfolio)'. (Wolf, 1999 cited in Barrett 2004, p. 1)

Wolf notes that these models can be at odds, philosophically, with each other. This observation suggests that when designing an e-portfolio, a clear focus on the purpose of the portfolio will be required.

Williams and Easingwood (2006) identify a number of issues for consideration if ICT is to be effectively incorporated into investigative science. They caution that when planning science activities that utilize the affordances of ICT, the key objectives of the lesson should be scientific ones.

...science in primary school should be largely practically-based and that ICT must be an integral part of the work. ICT can be used at different times during a scientific enquiry – it can be used for research, collecting data, analyzing information, recording findings and displaying and presenting the results...

Williams and Easingwood (2006, p. 9)

They also recommend that when learning to use databases it is better that children input and use their own data. Harlen and Qualter (2004, p.224) believe that the incorporation of ICT into investigative work is 'helping to bring a better balance to practical work' by enabling children to avoid the 'tedium' of making a written record of each aspect of the investigation.

A number of principles for the designing of an effective science enquiry assessment task emerge from this brief review of literature. Summative assessment tasks should:

- be achieved partially through carefully designed special tasks;
- be moderated with other teachers;
- framed in a way that allows children to succeed;
- not be seen as 'high stakes' or treated in isolation from other assessment data.

In particular, tasks should incorporate opportunities for children to:

- raise investigable questions, with support;
- hypothesise and predict;
- share and discuss ideas;
- design an investigation and select appropriate resources;
- collect evidence through practical investigations;
- reflect on their learning.

Further design principles for effective e-portfolio-based assessment emerge from this discussion. An e-portfolio is likely to offer:

- a variety of tools that allow children to capture their thinking, obtain and present data in an effective 'user friendly' way;
- a framework to encourage engagement with the full range of investigative skills and processes;
- an enhanced experience of investigative science and technology that allows children to demonstrate their abilities fully.

It was with these principles in mind that the research described below was conducted, and the assessment activities were designed.

## **RESEARCH QUESTIONS AND METHODOLOGY**

This study trialled and evaluated the e-scape approach to assessing children's procedural and conceptual understanding of science with 9 – 11 year-olds. The e-scape system itself consisted of an online task-authoring tool (MAPS 3, developed by TAG Developments); an exam-management system EMS) running on a laptop server in the classroom and connected wirelessly to a class set of 'Fizzbook Spin' netbooks. We used the system with primary teachers for three main purposes: (1) to develop tasks that were designed to engage pupils in stimulating enquiry, (2) to manage the running of the tasks in the classroom, and (3) to facilitate the assessment of e-portfolios. The netbooks that pupils used to record their work offered a range of multimodal responses (audio, text, video, still images, drawing) and had touch-sensitive screens (figure 1).



Figure 1: Fizzbook Spin netbook



There were two overarching aims of the project: firstly to develop an approach to e-portfolio assessment of scientific and technological understanding at age 11 and secondly to see how well the e-scape system facilitated this. During the course of the project, insights into a series of research areas were sought and these areas were used to make an evaluative judgement on the effectiveness of the system. Specifically judgements were made in five areas: (1) the reliability of running the e-scape system in the primary classroom; (2) the extent to which 9-11 –year olds found the e-scape system a stimulating way to engage with scientific enquiry; (3) comparison between the e-scape system and paper-based approaches to recording pupil responses; (4) the reliability and validity of assessments of pupil portfolios made by project participants using the e-scape ‘pairs comparison’ approach; and (5) the usefulness of the e-scape MAPS software in making criterion-referenced assessments. Data were collected by conducting classroom observations of the system in operation, interviewing teachers, reviewing e-portfolios generated by pupils, analysing statistics generated by the pairs comparison process and conducting a participant questionnaire.

Participants in the project were recruited from eight state primary schools in the south-west of England. A total of sixteen teachers and two trainee teachers participated in the study, plus 263 9-11

year-old pupils. The decision to invite two teachers from each school reflected the value we place on paired work. In a pair there is the opportunity for teachers to support each other in developing tasks and thinking, and it enhances the impact of the project in the school and beyond. The project consisted of five centrally-based training days spaced out over ten months with teacher participants required to carry out task development and trials in school between training days. Five of the participating schools were involved in the project from the start with another three joining the project from training day 3 onwards. The training provided and tasks carried out are summarised in table 2.

Table 2: Outline of project

Training Day 1	Teachers were introduced to the e-scape approach for assessing pupils' procedural and conceptual understanding, and the procedure for authoring e-scape tasks.
School-based Tasks and Trials	Participant teachers developed a paper-based version of an e-scape assessment task which was trialled in the classroom. Pupils explored the different ways that it was possible to record responses using the project's wireless laptops.
Training Day 2	Teachers were supported in transferring paper-based tasks into an electronic form using the authoring tool.
School-based Tasks and Trials	Teachers authored an electronic version of the paper-based task they had developed earlier and trialled this in the classroom.
Training Day 3	The e-scape approach to making judgments on pupils' work by paired comparisons was introduced. E-portfolios generated by pupils involved in the project were assessed using this approach.
School-based Tasks and Trials	Schools involved in the project from the start either refined a previously-authored task and trialled this in the classroom or devised a new task and trialled it. Newly-recruited schools authored and trialled paper-based and electronic versions of tasks based on the science topics they were covering
Training Day 4	A selection of e-portfolios generated by tasks carried out in schools were judged using pairs comparison and Assessing Pupil Progress in Science (APP) attainment statements (DCSF 2010). Participant teachers agreed on a task to be trialled in all schools that took as its focus the adhesive properties of sticky tapes.
School-based Tasks and Trials	The sticky tape task was jointly-authored online and then trialled in eight schools.
Training Day 5	Participant teachers assessed pupils' sticky tape e-portfolios using the pairs engine and APP criteria.

## FINDINGS

The findings from the project described here inform later discussions on the effectiveness of the e-scape system in assessing pupil enquiry skills at primary level. A diverse range of data contributed to

our evaluation of the system. Teacher interviews and their responses to questionnaires are described (the data has been amalgamated from nine teachers' responses), and the statistical reliability of the e-scape system's assessment tool is reported in this section. In total the project teachers produced 10 online assessment tasks covering the science topics of forces, electricity, materials, human body, micro-organisms, light and sound. Each involved an element of designing and, in some cases, making:

1. Bath Rugby Activity – First, pupils considered the physical attributes that a rugby player needs, they investigated some of these factors and finally 'designed' the ideal rugby player
2. Static Electricity – Factors that affect static electricity were investigated in response to an orientation activity that asked pupils to rub a balloon against their hair.
3. Shadows Activity – A video clip of shadow puppetry and a concept cartoon provided the stimulus for pupils to investigate shadow formation.
4. Changing Sounds – After exploring a collection of musical instruments, the factors affecting pitch of notes in one instrument were investigated further.
5. Paper Spinners – A range of factors that affect the time a paper spinner takes to fall were tested
6. Dissolving Task – A video clip of sugar dissolving in tea was the starting point for pupils to investigate the factors that affect the time it takes for sugar to go into solution.
7. Electrical Circuits – The brightness of lamps in different electrical circuits was the focus of this activity.
8. Electricity Investigation – A similar task to that above except the focus was specifically on changing the thickness of wire in the circuit.
9. Friction – factors that affect the grip of shoe soles were investigated.
10. Sticky Tape – the final task completed by all schools investigated the 'stickiness' or strength of a range of tapes. Figure 2 shows in detail how the task was structured. Note that the modes of response were varied (audio, video and still images, drawing, word processing, spreadsheets with graphing package) and that sometimes pupils had the option of choosing their preferred way of recording (figure 3). The task was authored in such a way as to scaffold the investigation for pupils while giving them independence in planning, measuring and evaluating, as well as recording.

Figure 2: Screen prompts from the sticky tape task in the order that they appeared during the investigation.

Test Student

Look at the collection of sticky tapes. Make a list of some of the DIFFERENCES between them. Type your answer here



Explain what these different sticky tapes could be used for, and why. You can EITHER record your voice OR type your answer OR draw a diagram

click here to add line

Test Student

Rank the tapes in order from stickiest to least sticky Take a photo of your rank order

Why do you think some are stickier than others? Try to explain your answer using scientific words like 'force'. You can EITHER record your voice OR type your answer OR draw a diagram.

click here to add line

Test Student

Think about something you would like to make or fix using sticky tape. Write it here.

Look at the resources available. How could you use these to test which tape to use? Draw a labelled diagram to show how you would use the resources to do your test.

Explain why you've chosen this method. You can EITHER Record your voice OR type your answer.

click here to add line

Test Student

Make a list of ALL the things you COULD CHANGE in your investigation here.

Choose ONE of these variables to investigate. Write it here.

Do you need to make this a fair test? Explain your answer by making a voice recording.

Make a list of the things you need to KEEP THE SAME here.

click here to add line

Test Student

What will you measure and how? Record your answer here.

How will you make your results as accurate as possible. EITHER record your voice OR type your answer.

click here to add line

Test Student

Click here to download a spreadsheet to your desktop. This will be your table of results. Open it and fill it in when you do your investigation.

Click here to upload your spreadsheet file When you have finished, use the spreadsheet graph tool to draw a graph. Drag the graph underneath the table and save the file.

Now make a short video to explain how you did your investigation and made your results accurate.

click here to add line

Test Student

Look back at your table and graph in the spreadsheet on your desktop. Are there any results that don't seem right? Record OR type your answer.

Are there any patterns in your results? Record OR type your answer.

So has your investigation helped you choose the right sticky tape for what you want to make or fix? Make a short video to explain your answer.

click here to add line

Test Student

How did working with a partner make your investigation better? EITHER record OR type your answer.

How could you improve your investigation? EITHER record OR type your answer.

click here to add line

A total of 284 online pupil e-portfolios for assessment across the tasks were created, including 154 e-portfolios across eight schools for the sticky tape task.

Figure 3: Pupil photographing 'rank order of tape stickiness' during 'Sticky stuff' task



### Responses to paper-based tasks

In order to help project teachers familiarise themselves with portfolio assessment of scientific enquiry, we initially asked them to run a paper-based activity in the e-scape format. The following excerpts from a teacher's oral account give a sense of the quality of pedagogy achieved. The teacher had created a paper-based task that asked pupils to investigate the human body in the context of designing a rugby player. The teacher commented that *'it was probably some of the best science I've done this year'*. He noted that *'some terrific discussion was generated which – had it been recorded - would have been the best evidence [of pupils' learning]'* and that he felt *'the discussion was really, really excellent – but it could have done with being recorded because it's now disappeared and it's only anecdotal'*. The latter comment underscored the value of the digital approach that was about to be launched. He made the points that in terms of learning *'it's the most open-ended I've been in science'* and that pupils made lots of mistakes, *'but that was where they were learning'*. He noted that it did take a lot of time to complete the task (*'it took us the whole afternoon'*) but pupils seemed unconcerned by this (*'they wanted to carry on the next day – and we did carry on next day and they spent another lesson writing up their results'*).

### **Teachers' comments on using the e-scape system**

At the point that the full digital e-scape technology was rolled out for use in the classroom there were some initial technical difficulties that made the teachers' task of using the system more challenging than was desirable. Despite this, a wide range of science and technology projects created by the teachers to support their class's science work were developed and trialled. In terms of their experience of using the authoring tool, teachers noted the following advantages:

*Task broken down made it easier for children*

*Lots of options for displaying info – movies-pics etc. The choice allowed teachers to present things in more than one way*

*Allowed you to select how you wanted things include*

*Ability to plan step-by-step boxes for the children to follow and (they) can choose how to present*

*Easy to set new task and create new boxes ..easy to assign whether you wanted children to speak/type/video etc*

Their comments about the benefits of scaffolding assessment through the step-wise e-scape process in the authoring tool suggest that some teachers had significantly developed their pedagogy. Other comments illustrate the impact of the e-scape approach on teachers' practice and are mainly focused on the multiple response mode possibilities for learners. This was seen as a strength in two main ways; as motivational, and as empowering for children with limited conventional writing skills:

*(I have) more awareness of and understanding in children who find it hard to write ideas.*

*Helps with playing towards child's strengths – choice of recording. Collecting a portfolio of child's work*

*Choice of recording is really supportive for children who struggle to record. Instant portfolio.*

*Fun for children .. good to use other forms of recording.*

*I think its fantastic and has implications for all other subjects. Its better as a source for collecting evidence currently.*

*Enjoyment of children ... multi-sensory.*

In terms of running the activity in the classroom using the netbooks, teachers focussed on the motivational nature of the Fizzbook Spins – and the potential this provides for enhancing learning:

*Children enjoyed using the fizzbooks ... robust, small and drawing mode (is good). ICT skills learned ... & independent working*

*Children enjoyed them .. good to have integrated system (video/camera/sound/drawing)*

*(Children with special educational needs) could access science/ICT. Portability, children's enjoyment. Durability.*

*Totally cool. Gives children an enthusiasm for the subject... they're excited and skilled to use them. Revolutionary.*

However, there were some practical difficulties with using the netbooks in practice in the classroom:

*Battery power! .. the question disappears as you type the answer. Some get distracted*

*Microphone placement (was poor)*

*Can be too small on screen*

*At times children spent more time playing with videos than focussing on investigation*

Some drawbacks of using the authoring tool were also noted:

*Need to be able to preview (the boxes) when writing – to see what children will see.*

*Took too long – and fiddly - need to think through on paper first.*

*Initially difficult to attach pictures / excel files – but this was sorted out.*

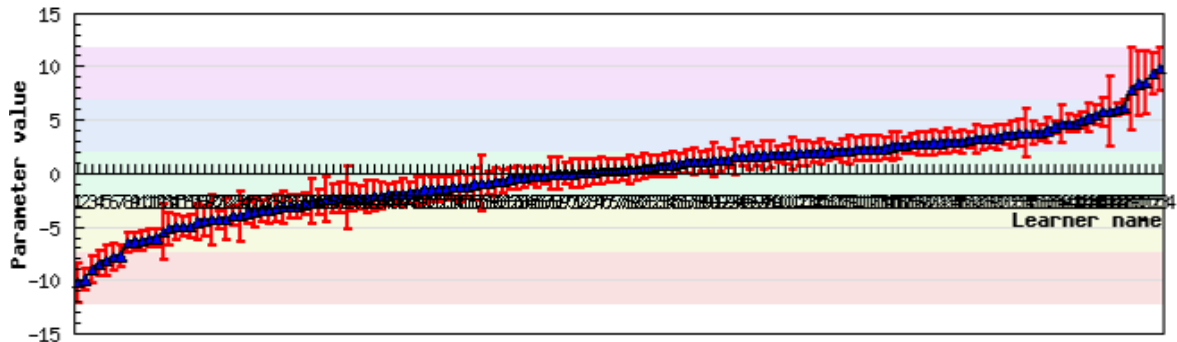
The e-portfolios produced were assessed using the e-scape system's "pairs engine", based on the Adaptive Comparative Judgement approach to assessment (see Pollit's article in this edition). Teachers' reflections on the process of designing an e-scape assessment task illustrated how they understood its iterative nature in that changes made to the activity inevitably had an impact on the portfolios produced. They noted, having worked through a series of comparative pair judgements, that *'task design would need to have clearer understanding of what aspects to be assessed'*.

### **Statistical reliability of using the e-scape 'pairs engine'**

The pairs judging process for the 154 'Sticky Stuff' e-portfolios demonstrated a high degree of reliability between the judgements made by the 17 teachers and tutors involved. 720 pairs judgements were made (an average of 42 judgements per judge) resulting in the rank order for the portfolios represented in figure 4. The process assigns each portfolio a parameter value, which is the likelihood of it 'winning' a pairs comparison judgement. Portfolios in the middle of the plot have a parameter value around zero, indicating that they have 'won' (or would 'win') comparisons with about half of the others.. Portfolios at the far left of the plot have parameter values approaching -10 indicating that they lose most of their comparisons, whilst those to the far right have parameter values approaching +10 as they tend to 'win' all their comparisons. The slightly longer error bars on some portfolios indicate a degree of disagreement between judges as to their position in the rank order. However,

overall this rank order has a reliability coefficient of 0.88 from just over 9 judgements per portfolio, which would be likely to rise to 0.9 or higher with further judging rounds.

**Figure 4: Parameter value error plot for pairs judging of 154 ‘sticky stuff’ portfolios**



## DISCUSSION

The above responses of teachers to e-scape portfolio assessment of primary scientific and technological understanding, together with the results from the pairs-assessment trial, raise the following issues:

### 1. Authenticity versus manageability of assessment activities

The e-scape authoring tool presents teachers with the opportunity to develop ‘special assessment tasks’ (Harlen 1999) to address the main elements of effective enquiry-based classroom practice (Harlen 2008, see literature review). For example, in developing the ‘design a rugby player’ task, the teacher was able to include:

- opportunities for children to collect evidence (around the physical attributes that might affect a rugby player’s performance);
- opportunities to express their ideas (about the parts of a rugby player’s body working the hardest in a photograph [interestingly many identified the brain]), to listen to the ideas of others (through working in pairs) and to build on their existing ideas when faced with new experiences (testing their ideas in the playground);
- opportunities to hypothesise (about which physical attributes might be more important), predict and suggest answers;
- prompts for discussing how to test their predictions (e.g. about leg length and running speed) and see if their ideas ‘work’;

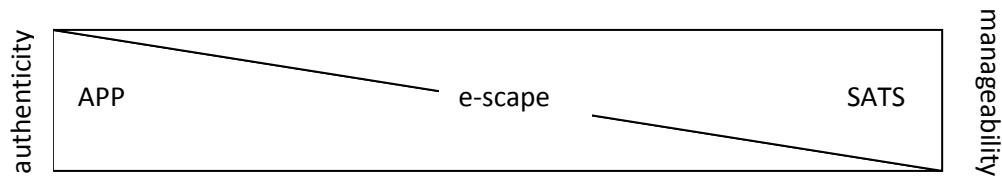


- a meaningful context so children were clear about what they were finding out and what they were learning by doing so;
- opportunities for children to consider the evidence they collected in relation to their initial ideas and predictions;
- a means, through the e-portfolio, to reflect and report on how and what they had learned.

In addressing the above criteria for meaningful enquiry, the task can claim greater authenticity than the widely-criticised and discontinued SATs tests (Harlen 2007). This authenticity as an assessment task did, however, come at the expense of manageability in the classroom. In particular, the task took much longer than anticipated for the children to complete – a feature noted by several of the teachers – since its open-ended nature required them to make decisions which subsequently proved to be mistakes. Although learning from these mistakes was valuable, the danger of children not completing tasks or not achieving as highly as they might do in a more directed activity was a source of anxiety for the teachers in the context of continuing high-stakes assessment and published school league-tables.

The government-recommended alternative to the e-scape approach – Assessment of Pupil Progress (APP) – is potentially less manageable for primary teachers. By suggesting that evidence is collected from pupils' everyday classroom activities to cover each statement within five assessment focuses (AFs) (DCSF 2010), the APP approach necessitates the accumulation and annotation of hundreds of pieces of work across a school year for a typical class of 30. Whilst there is arguably an authenticity in capturing pupil performance in the course of an activity which was not specifically designed as an assessment task, this means in practice that some criteria are covered repeatedly whilst others are extremely difficult to evidence. The review of APP pilots in English and Mathematics (QCDA 2008) claims a degree of manageability, but the science framework has not been piloted owing to a change of government. The responses of project teachers suggest that specifically-designed assessment tasks such as e-scape would be likely to be more manageable than the mass evidence accumulation required for APP science. If we conceptualise the trade-off between authenticity and manageability in assessment as in figure 5 – in which one declines as the other increases – the e-scape approach can be seen as lying between APP (high authenticity but low manageability) and the pencil-and-paper test approach of SATS (high manageability but low authenticity):

**Figure 5: Three assessment approaches positioned on a continuum representing the trade-off between authenticity and manageability in assessment**



Although e-scape tasks could be used in combination with either of the other approaches (Hodgson and Pyle 2010), even used on their own they represent a balanced compromise in relation to this continuing tension in summative assessment. It is worth adding that the authenticity referred to here is in relation to assessing ‘normal’ classroom activity; well-designed e-scape tasks have the potential to be highly authentic in the way in which they portray the activity of ‘doing science’.

## 2. Reliability of assessment judgements

Although there are different kinds of reliability in educational assessment (for example internal reliability of the task, reliability of judgement over time), perhaps one of the more significant measures is that of inter-marker reliability, since it would be unfair on children if the mark assigned to their work depended upon who marked it. If the consequences are high, such as they are in end-of-Key-Stage assessment in England, Wiliam (2000) argues that reliability needs to be high - 0.90 or above. The inter-marker reliability coefficient obtained in the Adaptive Comparative Judgement trial of the e-scape ‘sticky stuff’ activity – 0.88 – could therefore be regarded at nearly of a level to recommend its use for such purposes, and could easily be raised to this level by further judging rounds. It also compares reasonably well with reliability coefficients from SATs, which in the case of the 2001 tests ranged from 0.80 to 0.94 (Newton 2003). Although few studies of inter-marker reliability for teacher assessment have been conducted, where it has been compared with statutory tests (Reeves *et al.* 2001) there has been a tendency to under-rate boys’ performance in science at age 11, perhaps because some boys with secure scientific understanding and enquiry skills have difficulty in expressing these in writing during general classroom work. This suggests that e-scape may well be more reliable than teacher assessment, perhaps because the online assessment process provides for moderation across more markers than is normally possible for within-school moderation.

There is evidence to suggest that teachers should be able to score hands-on science investigations and projects – such as e-scape - with high reliability using detailed scoring criteria (Frederiksen & White, 2004), but the use of Thurstone-pairs marking (Pollitt & Crisp 2004, Greateorex *et al.* 2008) in this context is less well documented. Pollitt and Crisp (2004) present evidence that this method could lead to a more valid assessment by reducing the restrictions placed on the way that questions are written when the traditional marking is to be used, though it should be noted that the teachers in our sample recognised the need for task design to reflect a clearer understanding of what aspects were to be assessed. Although it is not criterion-referenced in the same way as APP, the ‘rank order’ generated can be converted to a level judgement. Teachers need to identify ‘grade boundaries’ within the overall sample: points above which all portfolios have met a particular set of criteria. For example, if we assume the sample to include work at levels 3, 4 and 5, teachers might start by looking at portfolios about one third of the way ‘down’ the rank order to see whether they fit the level 4 or 5 criteria best. Depending on the levels found, they would then look above or below this point until they could find a pair of portfolios, the lower of which meets level 4 criteria whilst the higher can be assessed at level 5. We can then assume that all the portfolios above this point can be awarded level 5, before going on to look for the level 3/4 boundary in the same way. This process may appear laborious, however in subsequent assessments of the same task, Meadows and Billington (2005) argue that it can be effectively side-stepped by including ‘grade marker portfolios’ from the previous year in the sample. Furthermore, the online nature of the marking process should also lend itself to greater reliability, since...

... e-marking allows more effective monitoring of examiner reliability while marking is underway, allowing the identification and investigation of problems at an early stage, when interventions can be made most efficiently (Meadows and Billington 2005, p. 67).

### **3. Affordances of ICT in the assessment process**

As well as potentially increasing inter-marker reliability, the ICT-rich nature of the e-scape system appears to have been particularly significant for the teachers, despite the inevitable technical challenges. There was evidence of motivation and empowerment, as well as support for children's investigative skills (Williams and Easingwood 2006) and their ability to communicate their learning (Murphy 2006, p.19) using the multimodal functionality of the Fizzbook netbook computers. In terms of Wolf's typology of e-portfolio use, the e-scape model could be seen as an assessment tool to document the attainment of standards (cited in Barrett 2004), although in places there were 'stories of deep learning' (ibid.) and several of the teachers saw how it could be used formatively within a constructivist model of assessment.

### **4. Impact of assessment on teaching**

The willingness of project teachers to incorporate the notion of e-portfolios into their assessment approaches and to make use of the authoring tool structure to scaffold enquiry tasks using action and reflection steps demonstrates a positive version of the 'backwash' effect of assessment into pedagogy and the curriculum (Harlen and Deakin-Crick 2002). Rather than reducing the amount of practical science in the classroom as observed by Wyse *et al.* (2008) in the case of national curriculum testing, the e-scape approach appears – at least to judge by teachers' expressed intentions – to prioritise practical enquiry, much as observed by Collins *et al.* (2008) (cited in Tymms *et al* 2010) following the abolition of statutory science testing at age 11 in Wales. Although some recognised the need to design tasks to better match the criteria they hoped to assess, other participants commented on how open-ended their e-scape tasks had made the science and technology going on in their classroom. Because of the changing political context during the project, the technological aspects of learning and assessment had become less central to the tasks for many teachers, but from the few examples where it had been a feature (designing guitars, selecting tape to make or mend an artefact) it was possible to see how this approach might facilitate subject integration.

## CONCLUSIONS AND RECOMMENDATIONS

The outcomes from our project suggest that an e-scape approach to the assessment of scientific enquiry has potential to be authentic and reliable, and that it may even have a positive effect on current pedagogy. A follow up study could explore this assertion, moving the project beyond the stage where novelty has a potential impact on outcomes. Such a study could explore the prediction that e-scape authored and delivered science enquiry activities will lead to more frequent opportunities for children to carry out scientific investigations that are relevant, engaging and challenging. In turn, this may impact positively upon children's attainment in scientific enquiry. E-scape requires a classroom to be equipped with devices with wireless connection to a server or the internet. While this equipment is not yet commonplace, it is not unusual to find it in a primary school. Alongside this investment in new technology, training and technical support for teachers would present an additional cost until the software becomes more sophisticated or 'user friendly'

The importance of peer and self-assessment in children's learning is highlighted by the Assessment Reform Group (2008) and Harlen and Qualter (2009) among many. Once an e-portfolio has been created it is in an ideal form to be reviewed by the learner and classmates. The e-scape system allows for both self- and peer-assessment, but although pupils were invited to review their e-portfolios at the end of each session, this was not a specific focus of our project and would also merit further research. As noted above, e-portfolios can take a number of forms for philosophically different purposes (Wolf 1999, cited in Barrett (2004)). The portfolios generated during this project, based on single activities, could not be described as 'digital stories of deep learning', although if e-scape were to be imbedded in classroom practice and children habitually recorded their actions, thoughts and reflections during day-to-day activities, the resultant portfolios could indeed be comprehensive and 'deep' multifaceted records of learning. Alternatively, as at present, e-scape e-portfolios can provide an accessible record of attainment, able to be norm or criterion referenced, internally and externally moderated. Further consideration needs to be given to this potential dichotomy of purpose, since if the learner is aware that judgements are to be made on the portfolio contents, this may impact upon their willingness to represent their 'authentic' learning through it.

## ACKNOWLEDGEMENT

The authors wish to thank the AstraZeneca Science Teaching Trust for their financial support for this research.

## REFERENCES

- Assessment Reform Group (2008) *Changing Assessment: Practice, Process, Principles and Standards*. Online. Available < <http://www.assessment-reform-group.org/ARIA%20English.pdf> > Accessed 28 01 11
- Barrett, H. (2004) *Electronic Portfolios as Digital Stories of Deep Learning*. Online. Available <<http://electronicportfolios.org/digistory/epstory.html>> Accessed 07 12 10
- Davies, D. (2009) Digital Portfolio Assessment of Secondary Students' Scientific Enquiry Skills: the e-scape project, *European Science Education Research Association (ESERA) 2009*, Istanbul, Aug 31 - Sept 4.
- Department for Children, Schools and Families (DCSF) (2010) *Assessing Pupil Progress (APP) Overview*.  
<http://nationalstrategies.standards.dcsf.gov.uk/primary/assessment/assessingpupilsprogressapp>. Accessed 24.1.11.
- Frederiksen, J. & White, B. (2004), Designing assessment for instruction and accountability: an application of validity theory to assessing scientific inquiry, in M. Wilson (Ed.) *Towards coherence between classroom assessment and accountability, 103rd Yearbook of the National Society for the Study of Education Part II* (Chicago, National Society for the Study of Education), 74–104.
- Gillespie, H. (2006) 'Virtual learning in primary science' in Warwick, P. Wilson, E. Winterbottom, M. (eds), *Teaching and Learning Primary science with ICT*. Maidenhead: OUP.
- Greatorex, J., Novakovic, N. & Suto, I. (2008), What attracts judges' attention? A comparison of three grading methods, *Annual Conference of the International Association for Educational Assessment*, Cambridge, September 2008.
- Harlen, W. (1999). 'Purposes and procedures for assessing science process skills', *Assessment in Education*, 6, 1, 129–144.
- Harlen, W. (2007) *The Quality of Learning: assessment alternatives for primary education*. (Primary Review Research Survey 3/4), Cambridge: University of Cambridge Faculty of Education. ISBN 978-1-906478-03-2.
- Harlen W. (2008) Science as a key component of the primary curriculum: a rationale with policy implications. *Perspectives on Education 1 (Primary Science)*, 2008:4–18. Online. Available at [www.wellcome.ac.uk/perspectives](http://www.wellcome.ac.uk/perspectives) [accessed 2 Dec 2010].
- Harlen W, Deakin Crick R (2002). A systematic review of the impact of summative assessment and tests on students' motivation for learning (EPPI-Centre Review, version 1.1). In: *Research Evidence in Education Library*. Issue 1. London: EPPI-Centre, Social Science Research Unit, Institute of Education.
- Harlen, W. and Qualter, A. (2004) *The Teaching of Science in Primary Schools* (4<sup>th</sup> edn.) London: David Fulton
- Harlen, W. and Qualter, A. (2009) *The Teaching of Science in Primary Schools* (5<sup>th</sup> edn.) London: David Fulton

- He, Q., Hayes, M. and Wiliam, D. (2011). *Classification accuracy in results from Key Stage 2 National Curriculum tests*. Coventry, Ofqual.
- Hodgson, C. and Pyle, K. (2010) *A Literature Review of Assessment for Learning in Science*. Slough: NFER. Online. Available [http://www.nfer.ac.uk/nfer/publications/AAS01/AAS01\\_home.cfm?publicationID=436&title=literature%20review%20of%20Assessment%20for%20Learning%20in%20science](http://www.nfer.ac.uk/nfer/publications/AAS01/AAS01_home.cfm?publicationID=436&title=literature%20review%20of%20Assessment%20for%20Learning%20in%20science) [Accessed 07 12 10].
- Howe, A., Davies, D. McMahon, K., Towler, L., Collier, C. and Scott, T. (2009) *Science 5-11: A Guide for Teachers* (2<sup>nd</sup> edn.) London: David Fulton
- Kimbell, R., Wheeler, A., Miller, S. & Pollitt, A. (2007) *E-scape portfolio assessment phase 2 report*, Department of Design, Goldsmiths, University of London.
- Kimbell, R., Wheeler, A., Miller, S. & Pollitt, A. (2009) *E-scape portfolio assessment phase 3 report*, Department of Design, Goldsmiths, University of London.
- McLaren, S. (2010) *Assessment is for Learning through digital technologies [e-scape scotland]* , available online: [http://www.sqa.org.uk/files\\_ccc/Summary%20Research%20and%20Development%20Report.pdf](http://www.sqa.org.uk/files_ccc/Summary%20Research%20and%20Development%20Report.pdf) [accessed 16.9.11].
- Meadows, M. and Billington, L. (2005) *A Review Of The Literature On Marking Reliability*. London: National Assessment Agency. [http://orderline.qcda.gov.uk/gempdf/184962531X/QCDA104983\\_review\\_of\\_the\\_literature\\_on\\_marking\\_reliability.pdf](http://orderline.qcda.gov.uk/gempdf/184962531X/QCDA104983_review_of_the_literature_on_marking_reliability.pdf) [Accessed 24.1.11].
- Murphy, C. (2006) 'The impact of ICT on primary science' in Warwick, P. Wilson, E. Winterbottom, M. (eds), *Teaching and Learning Primary Science with ICT*. Maidenhead: OUP.
- Newton, P. (2003) The defensibility of national curriculum assessment in England. *Research Papers in Education*. 18(2), 101-126.
- Pollitt, A., & Crisp, V. (2004) *Could comparative judgements of script quality replace traditional marking and improve the validity of exam questions?* Paper presented at the BERA Annual Conference, UMIST Manchester, September 2004.
- Qualifications, Curriculum and Development Agency (QCDA) (2008) *Evaluation of the assessing pupils' progress in key stage 2 pilot project, 2006–2008*. Coventry: QCDA.
- QCDA (2010) *Assessing pupils' progress: learners at the heart of assessment*. Coventry: QCDA. [www.qcda.gov.uk/resources/assets/Assess\\_pupils\\_progress\\_webop.pdf](http://www.qcda.gov.uk/resources/assets/Assess_pupils_progress_webop.pdf) [Accessed 13.1.11].
- QCDA (2010) *National Curriculum Science Key Stage 2*. Online. Available < <http://curriculum.qcda.gov.uk/key-stages-1-and-2/subjects/science/keystage2/index.aspx> > [Accessed 07 12 10].
- Regional Support Centre Scotland (2009) *Assessment is for Learning through Digital Technologies: e-Scape Scotland*. Online. Available < <http://scottish-rscs.org.uk/newsfeed/wp-content/uploads/2009/09/escape.pdf> > [Accessed 07 12 10].
- Reeves, D. J., Boyle, W. F. & Christie, T. (2001) The relationship between teacher assessment and pupil attainments in standard test/tasks at Key Stage 2, 1996–1998, *British Educational Research Journal*, 27, 141–160.
- Rose, J. (ed.) (2009) *Independent Review of the Primary Curriculum: Final report*, Nottingham: DCSF Publications.

Tymms, P., Bolden, D. and Merrell, C. (2010) *Science in English primary schools: trends in attainment, attitudes and approaches Perspectives on Education 1 (Primary Science)*, 2008:19-41. Online. Available at [www.wellcome.ac.uk/perspectives](http://www.wellcome.ac.uk/perspectives) [accessed 2 Dec 2010].

Ward, H. Roden, J., Hewlett, C. and Foreman, J. (2005) *Teaching Science in the Primary Classroom*. London: Paul Chapman Publishing.

Williams, J. and Easingwood, N (2006) 'Possibilities and practicalities: planning, teaching and learning science with ICT' in Warwick, P. Wilson, E. Winterbottom, M. (eds), *Teaching and Learning Primary Science with ICT*. Maidenhead: OUP.

William, D. (2000) Reliability, validity, and all that jazz. *Education*, 29(3) pp. 9-13.

Wyse, D., McCreery, E. and Torrance, H. (2008) *The Trajectory and Impact of National Reform: curriculum and assessment in English primary schools* (Primary Review Research Survey 3/2), Cambridge: University of Cambridge Faculty of Education.